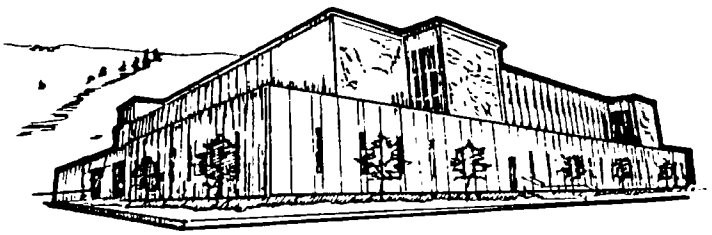
1993

Thermal desorption spectroscopy study of the initial oxidation of Si(111)

Joseph Rhea Gladden
The University of Montana

JOSEPH R GLADDEN



Maureen and Mike MANSFIELD LIBRARY

The University of Montana

Thermal Desorption Spectroscopy Study of the

Initial Oxidation of Si(111)

by

Joseph Rhea Gladden III

B.S. University of the South, Sewanee, TN, 1991

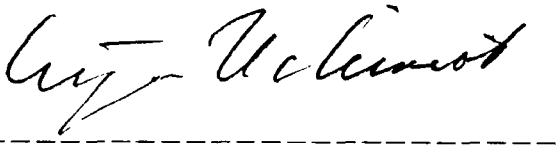
Presented in partial fulfillment of the requirements for the degree of

Master of Science

University of Montana

1993

Approved by:



Chairman, Board of Examiners

Dean, Graduate School

Dec. 8, 1993

Date

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Gladden, Joseph Rhea, III M.S., December 1993 Physics

Thermal Desorption Spectroscopy Study of the Initial Oxidation of Si(111)

Committee Chairman: Eijiro Uchimoto E. U.

Hydrogen terminated silicon has generated a great deal of interest because of its ability to protect the silicon surface against the formation of native oxides prior to thermal oxidation. This thesis reports the results of thermal desorption spectroscopy experiments designed to help elucidate the relationship between hydrogen termination and oxide growth retardation. The first part of this thesis deals with the problem of creating an atomically smooth Si(111) surface with dangling bonds terminated by deuterium. We have found that a process of dilute DF(1.6% in D_2O) treatment for 10 minutes followed by boiling in very dilute DF(0.005%) for 10 minutes creates an atomically smooth and oxide free deuterium (D) terminated Si(111) surface. The second part is a study of the initial stages of oxidation of the D terminated Si(111). Specifically, we have observed the dependence of SiO and deuterium surface coverage on the ambient temperature of the oxidizing environment, with a temperature range of 22 to 350°C. We found that the SiO TDS areal peak intensity increases linearly with oxidizing temperature for both the smooth and rough surfaces. Both show a SiO peak intensity increase of the same slope when plotted as a function of oxidizing environment temperature, but the smooth surface data is shifted about 50° C higher than the rough surface data. This indicates that the smooth surface causes a delay in the onset of oxidation. We have also found that the D_2 peak intensity decreases exponentially with temperature in the temperature range observed for both smooth and rough surfaces with the smooth surface data shifted about 30°C higher than the rough surface data. Finally, we believe that the difference in temperature dependence of SiO and D_2 data and the difference in the threshold temperature between smooth and rough surface data supports the model in which the early stages of oxidation does not occur by a simple exchange of the terminating D for O species, but that the O is initially inserted in the Si backbond between the first and second monolayers of the crystal.

Table of Contents

List of Figures

Acknowledgments

There are a great many people who provided assistance and support during the research for and writing of this thesis. I would like to thank some of them here. First, I would like to thank the NEC corporation which gave me the opportunity to spend six months in Japan on an educational internship during which time I was able to complete the experimental part of this thesis as well as learn a great deal of the Japanese culture and experience the hospitality there first hand. In particular, I would like to thank Mr. Yasushi Kaito who made this internship possible.

I would also like to thank Dr. Iwao Nishiyama of NEC who provided the guidance I needed in the design of the experiments and the interpretation of their results. Without his support, this thesis could never have been completed. Along the same lines, I would like to express my gratitude to my thesis advisor, Dr. Eijiro (Ebo) Uchimoto, whose dedication and hard work are appreciated by all of his students. He devoted much of his valuable time to helping me put this thesis together. The other members of my thesis committee, Dr. Randy Jeppesen and Dr. Douglas Klarup, also certainly deserve thanks for their well thought out comments and suggestions.

Lastly, I would like to show my appreciation to my parents Joe and Sally Gladden. Their moral, emotional, and financial support was invaluable during four years of college, two years of graduate school, and finally the completion of this thesis.

I would like to dedicate this thesis to my great-grandfather Dr. Woolford Baker who progressed from his birth on a cattle drive in what is now the state of New Mexico to receive a doctorate from Columbia University and become a professor of Biology at Emory University in Atlanta, GA from 1925 to 1961. He still lives in Atlanta at the age of 101 and continues to pass down a curiosity for the world to the generations that follow him.

Chapter 1: Introduction

1-1 Transistors - Historical Background

The first documented thoughts on the architecture and use of a metal-oxide-semiconductor (MOS) transistor were contrived by a Polish immigrant named Julius Edgar Lilienfield in New York City.[1] His designs received United States patents in 1930, 1932, and 1933. Development of the first working transistor was delayed by World War II, but Lilienfield's basic concepts helped pave the way for William Shockley of Bell Telephone Labs to design the first modern transistor. In December, 1947, he observed a voltage amplification of 2, and a power amplification of 330. By early 1948, a current amplification of 1000 was achieved. Shockley's invention was the first bipolar transistor and was probably the first to satisfy the current IEEE definition of a transistor which is a solid-state electron device with three or more terminals that has a power gain of greater than unity. John Bardeen and Walter Brattain, also of Bell Labs, announced the development of the point contact transistor shortly after Shockley's announcement. Bardeen, Brattain, and Shockley won the Nobel prize in physics for their efforts in 1948.[2]

The impact of these developments on the modern world is so great as to be hard to measure. These small solid-state devices which can amplify signal strengths began to replace the much larger and less reliable vacuum tubes in electronic devices such as radios, televisions, and eventually computers during the 1960's. With the advent of the transistor, the

electronic age, the necessary precursor to the space and information age, had officially begun. However, the road from the first working transistors to the present day computers and advanced consumer electronics such as digital wristwatches that can tell one not only the time but temperature, barometric pressure, and elevation trends was not easy or quick. It took the effort and dedication of literally thousands of scientists, doing fundamental research, and electrical engineers, learning how to put all the bits and pieces together, to make it possible. Some of these developments will be outlined in the following pages.

The first bipolar junction transistor was built by the team of Teal, Spark, and Buehlor of Bell Labs in 1951.[3] They used germanium as the semiconductor and made the junction by alternatively using n-type and p-type dopants during the growth of the crystal.*(see "Technical Introduction")* About one year later, Shockley, Read, and Hall began to study the rate at which electron and hole pairs are generated and recombined in the semiconductor material.[4] Their results, known as the SRH recombination kinetics (or statistics), are still used today and was an important stepping stone for the development of the dynamic random access memory (DRAM) cell.

One of the problems with the MOS transistors at this time was that they were very unstable electrically. This instability manifested itself by way of leakage current and low frequency noise due the trapping of electrons and holes in interface traps at the semiconductor surface. The period from 1958 to 1962 was a time of intensive research into the stabilization of the MOS transistor. The first major break through can be attributed to a team led by Atalla working for Bell labs. In 1959 they reported that by growing a thin

layer (150 to 300Å) of silicon dioxide on the silicon substrate surface, they greatly increased the stability of the metal-oxide-semiconductor field effect transistor (MOSFET).[5] They reported a reduction in the reverse leakage current by a factor of ten to one hundred, as well as a large reduction in the low frequency noise. Many people in the field saw this discovery as the stepping stone that moved the MOSFET from pure science and academic interest into the realm of applied technology. It paved the way for the first silicon integrated circuit and, later, volume production of MOSFET devices. Their method of growing thin oxide films on the silicon surface at temperatures of around 1000°C became known as thermal oxidation and is still used today in the manufacture of integrated logic circuits and RAM chips.

Several years after Atalla's passivation technique was reported, another major development came to light. This was the conception and design of the floating-gate MOSFET used as a memory device. In 1961, Sah observed that charge can be injected and stored for several days on the gate electrode of the bipolar junction transistor (**BJT**). A year later, Wanless of Fairchild laboratory recognized that Sah's observation meant the BJT could be used as a memory device. At the time, the major effort was still perfecting the silicon passivation techniques, so it was not until almost ten years later, when Frohman-Bentchkowski left Fairchild and joined Intel, that the idea was fully developed into a commercial product.

The next major milestone in the history of the MOSFET was the invention of the MOS integrated circuit. The basic idea of an integrated circuit is to put a large array of MOSFETs on one chip in order to serve as a memory device or form a logic circuit. The number of MOSFETs on the chip

designates it's "logic" power. During the period of 1969 to the present day we have progressed from small-scale integration (SSI) which consisted of about 100 MOSFETs to the present very large scale integration (VLSI) which provides about 100,000 MOSFETs per chip. The current cutting edge in computer chip technology is ultra large scale integration (ULSI) which will give us chips that contain more than one million MOSFETs per chip. This dramatic increase in the number of MOSFETs contained in one chip would seem to indicate an equally dramatic increase in the size of the chips. However, equally impressive advancements in the miniaturization of the MOSFETs and production technique during this time has kept the increase in chip area down to a factor of four.

The first product which made use of these transistors integrated into a circuit was 256 bit random access memory (RAM) chip introduced by INTEL in 1969. IBM soon followed suit with their 512 bit chip. A significant advance was made by Dennard of IBM in 1973 when he developed the scaling law which allowed the transistor dimensions to be reduced without affecting it's performance.[6] The bit count on the chips increased rapidly to 1k in 1970, 16k in 1976, 64k in 1979, on up to the current RAM chips which are on the order of 64 million bits per chip. Achieving these higher and higher bit counts requires that the MOSFETs become smaller. Keeping the chips small not only allows more chips to be packed in a smaller space, but actually makes them faster because electrons have a shorter distance to travel during charging (writing) and discharging (reading). It is also more efficient to manufacture because you can get a higher yield from the silicon wafer substrate on which the circuits are built. In order to further reduce the size of the MOSFETs, several obstacles need to be overcome. One of which is

that the oxide layer must be made thinner. As shown in Fig. 1, in the original 256 bit chip made by INTEL, the MOSFETs had an oxide thickness of 1500Å while the current generation of chips typically have a thickness of 150Å.

Date	Memory Capacity (bits)	SiO_2 Thickness (Å)
1965	64	1500
1976	16k	700
1979	64k	400
1987	4M	150

Fig. 1: A table showing the decreasing thickness of the oxide layer of progressively advancing RAM chips.

It is for this reason that there is still a concentrated effort in the industry to learn more about the silicon oxidation kinetics and why I have chosen this subject for my thesis. The next generation of chips will require even greater bit density which means that thinner and thinner oxidation layers must be reliably grown.

This work is primarily a study of the initial silicon oxidation kinetics. The oxide layer thicknesses dealt with in this thesis are in the range of 0 to 15Å and are therefore significantly thinner than those that are currently grown on integrated circuit chips. Hopefully, this research will give some insight as to how the first few monolayers of oxygen adsorb onto the silicon surface and whether layers this thin can be used in the manufacture of smaller MOSFETs and integrated circuit chips.

1-2 Transistors Technical Properties

The preceding section has given a brief historical overview of the development of transistors. It is the goal of this section to describe some of

the transistor's technical features such as basic theory, design, and applications. As stated above, the definition of a transistor used by the IEEE and Bell Telephone Laboratories is a solid-state electron device which has three or more terminals and a power gain of greater than unity. This is a vague definition at best and covers a great number of semiconductor devices. A comprehensive explanation of all these devices is beyond the scope of this thesis, but a brief review of some of the most common current transistors, including the bipolar junction (BJT) and MOSFET, as well as an introduction to semiconductors, is given below.

Before any transistors can be properly described, some properties and terms regarding semiconductors must be presented. A very simple definition of a semiconductor is a material that has electrical conduction properties somewhere in between a conductor, such as copper, and an insulator, such as plastic. To be more precise, a semiconductor's atoms have an energy gap between the valence and conduction bands equal to or smaller than about 1 electron-volt (eV) which is small compared to that of an insulator. To better explain this distinction, the concept of the Fermi energy must be introduced.

The Fermi-Dirac distribution function gives the probability of finding a fermion, such as an electron, in a certain energy state E at a particular temperature T, and is expressed by the formula

$$f(E) = \frac{1}{e^{(E-E_f)/kT} + 1}$$

where E_f is a constant known as the Fermi energy and k is Boltzmann's constant. Hence the Fermi energy level is the energy level at which the probability of finding an electron is exactly 1/2. Since f(E) decreases from 1 to

0 monotonically with increasing energy , we can see that it is more probable to find electrons in energy levels below the Fermi level than above it. At absolute zero (T=0), f(E) will be a step function with a value of 1 for $E<E_f$ and 0 for $E>E_f$.[7] In simple terms, the Fermi energy level can be thought of as a line dividing energy regions that contain most of the electrons and those that contain fewer electrons. The conduction energy band, which contains those electrons free to move if an electric field is applied, lies above the Fermi level while the valence band lies below. However, the electrons are not fixed to any one energy band and can cross the Fermi level if energy, thermal for instance, is either added to or extracted from the atom.

As stated above, a semiconductor has a small energy separation between the conduction and valence bands. Therefore, it can be easy for electrons to pass from one to the other if the conditions are right. In some situations, such as high temperatures, the material can act as a conductor and insulators in others. It should also be noted here that semiconductors have a crystalline structure, so its atoms are joined in a regular array by shared valence electrons in a covalent bond. This means that if an electron is excited from the valence to the conduction band, it leaves behind a space that can then be occupied by a valence electron from a neighboring atom. This space, or hole as it is usually referred to, acts like a positive charge and will move in the direction of an electric field if one is applied. So, when an electron jumps across the Fermi level, it is said that an electron-hole pair has been created.

If there is no contamination in the semiconductor, that is it contains only one type of atom, silicon or germanium for example, then there are an equal number of electrons and holes and it is said to be an intrinsic

semiconductor. However, if trace amounts of different atoms are introduced into the crystal structure, the electrical properties can be changed. Two types of dopants, foreign atoms, can be used, those that donate electrons and those that donate holes. The first type has more valence electrons than there are bond sites in the crystal structure, so one or more of its electrons is left free to combine with a hole elsewhere. The result is a surplus of electrons and an n -type semiconductor. The second type of dopant has fewer valence electrons than bond sites, so it absorbs electrons from its neighbors and results in a surplus of holes or a p -type semiconductor. This process is known as doping the semiconductor and is a vital step in the construction of many devices.

The IEEE definition above states that a transistor must have at least three leads, but says nothing about how those leads come together inside the device. Figures 2 and 3 show a schematic of the basic construction of the junction transistor and its circuit symbol, respectively.[8] As can be seen, it consists of three major components, the emitter and collector on the outside with the base, which is comparatively narrow, sandwiched between the two. There are two basic types of junction transistors depending on how each of the components are doped, the *pnp* and the *npn* . They both have the same basic construction and have the same operational characteristics, so only the *pnp* type will be described here.

The *pnp* junction transistor gets its name from the fact that the emitter and collector are p doped while the base is n doped. If a voltage is

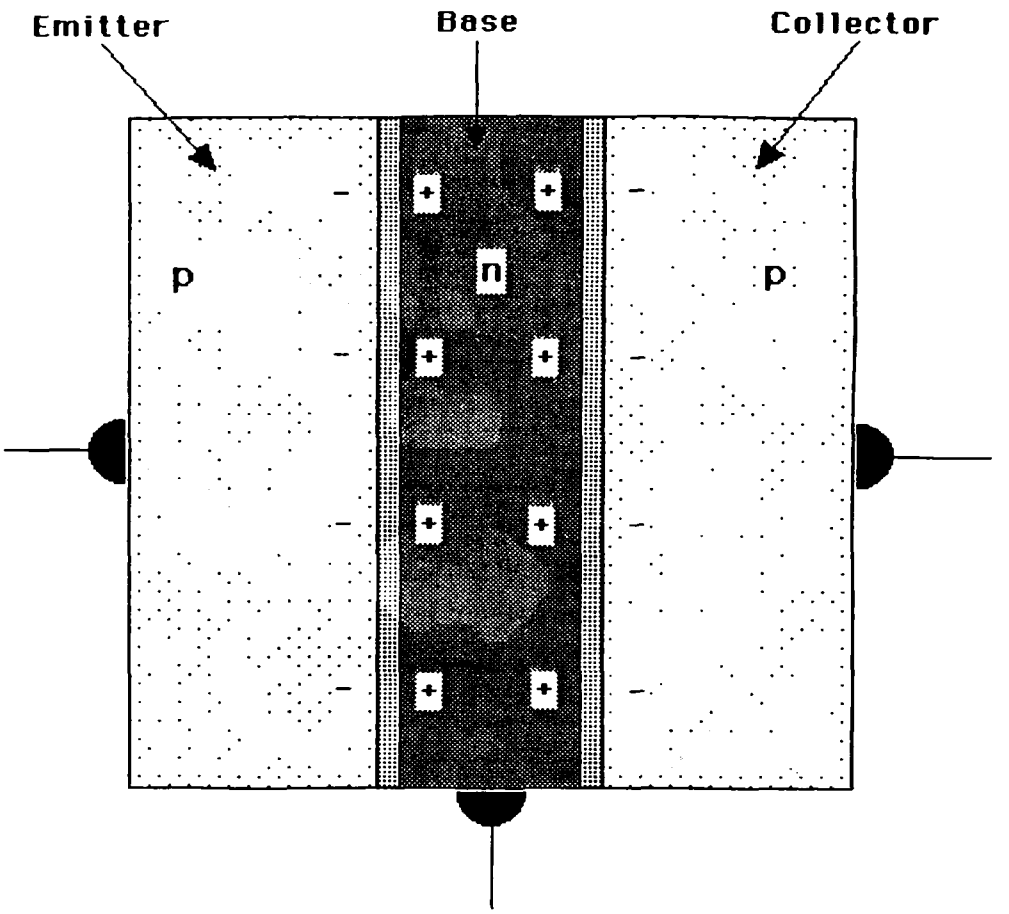


Figure 2: A schematic diagram of the *pnp* junction transistor.

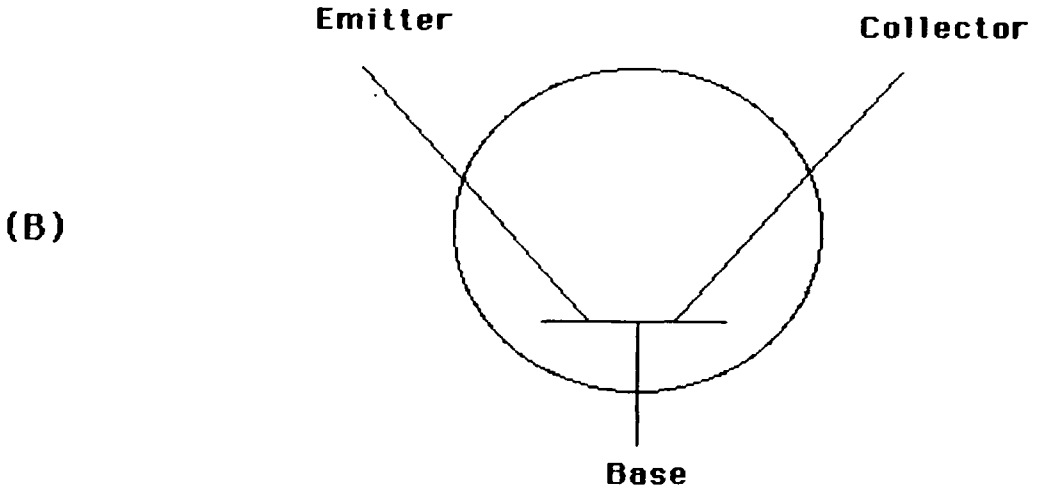


Figure 3: The circuit symbol for a transistor.

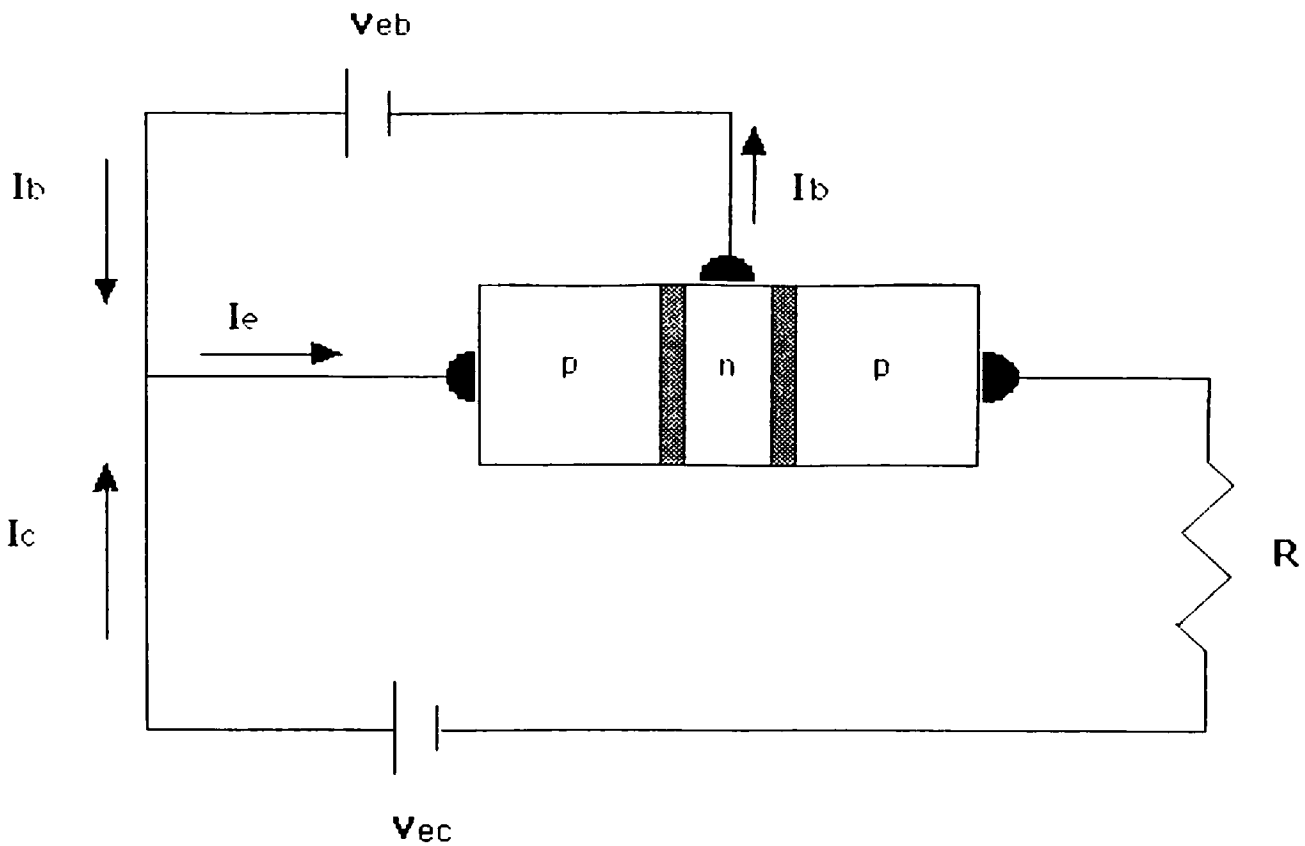


Figure 4: A *pnp* junction transistor hooked up in a circuit.

applied across the transistor, as shown by V_{ec} in Fig. 4, it will drive a current in the direction of the emitter toward the collector.[9] This condition creates a forward biased junction between the emitter and base and a reversed biased junction between the base and collector. The emitter is heavily *p* doped so there is a large surplus of holes. Many of these holes are forced into the base because of the forward biased junction. The base is medium *n* doped, but very narrow compared to the outer regions. The result is that some of the holes recombine with the excess electrons in the base, but the vast majority will cross this region, because it is so narrow, and will be pushed into the collector by the electric field at that junction. This occurs despite the fact that the base-collector junction is reversed biased. This is an important feature of the junction transistor.[10]

While most of the holes do cross the base, some recombine with the electrons there. This acts as an accumulation of positive charge in the base and limits the current flow. To prevent this from happening, an additional voltage source is applied across the base and emitter as shown in Fig. 4 as V_{eb}. This potential draws off the holes that do not make it through the base to the collector. The current flowing out of the collector I_c is very sensitive to the current coming out of the base I_b, which can be adjusted by varying V_{eb}. Small changes in I_b produce large changes in I_c which makes junction transistors ideal for time varying signal amplification. For such a use, one would hook up the varying voltage to be amplified as V_{eb} which would produce a small alternating current (ac), and a larger dc voltage source as V_{ec}. The ac will produce a time varying concentration of positive charges in the base which will in turn produce a large changes in I_c. This could then be measured as large voltage changes across the resistor R. This is perhaps the

most widely used application of the bipolar junction transistor (BJT) and the reason it is so useful in electronic devices which receive weak signals that must be amplified such as televisions and radios.

The surface field effect transistor (FET) works on a very different physical principle from the BJT and is the base with which modern computers are built. A cross section of a typical FET is shown in Fig. 5.[11] This is an n -channel device which consists of a p doped semiconductor as the base or substrate with small n type channels at one surface. Electrical leads are attached directly on top of the n region and are known as the source and drain for reasons that will soon become apparent. A larger lead is attached to an insulating layer, usually silicon dioxide, which is grown on top of the p substrate. This connection is known as the gate. Finally a lead is attached to the bottom of the substrate and grounded. It should be noted that the source, gate, and drain are all separated by the insulating SiO_2 layer. These types of devices are also referred to as metal-oxide-semiconductor field effect transistors or MOSFETs.

The FET uses an electric field to induce and vary a current between the source and drain. This is achieved by applying a large voltage to the gate which produces an electric field pointing from the gate to the grounded lead on the bottom of the substrate. This field forces more electrons into the substrate region directly below the gate and between the n type channels. However, since the SiO_2 layer has insulating properties, the electrons will be trapped in this region. If the applied voltage is large enough, the number of electrons in this region will begin to outnumber the number of holes. So, in effect, this area has been changed from a p type semiconductor into an n type and is known as an inversion layer. This creates a pathway for current

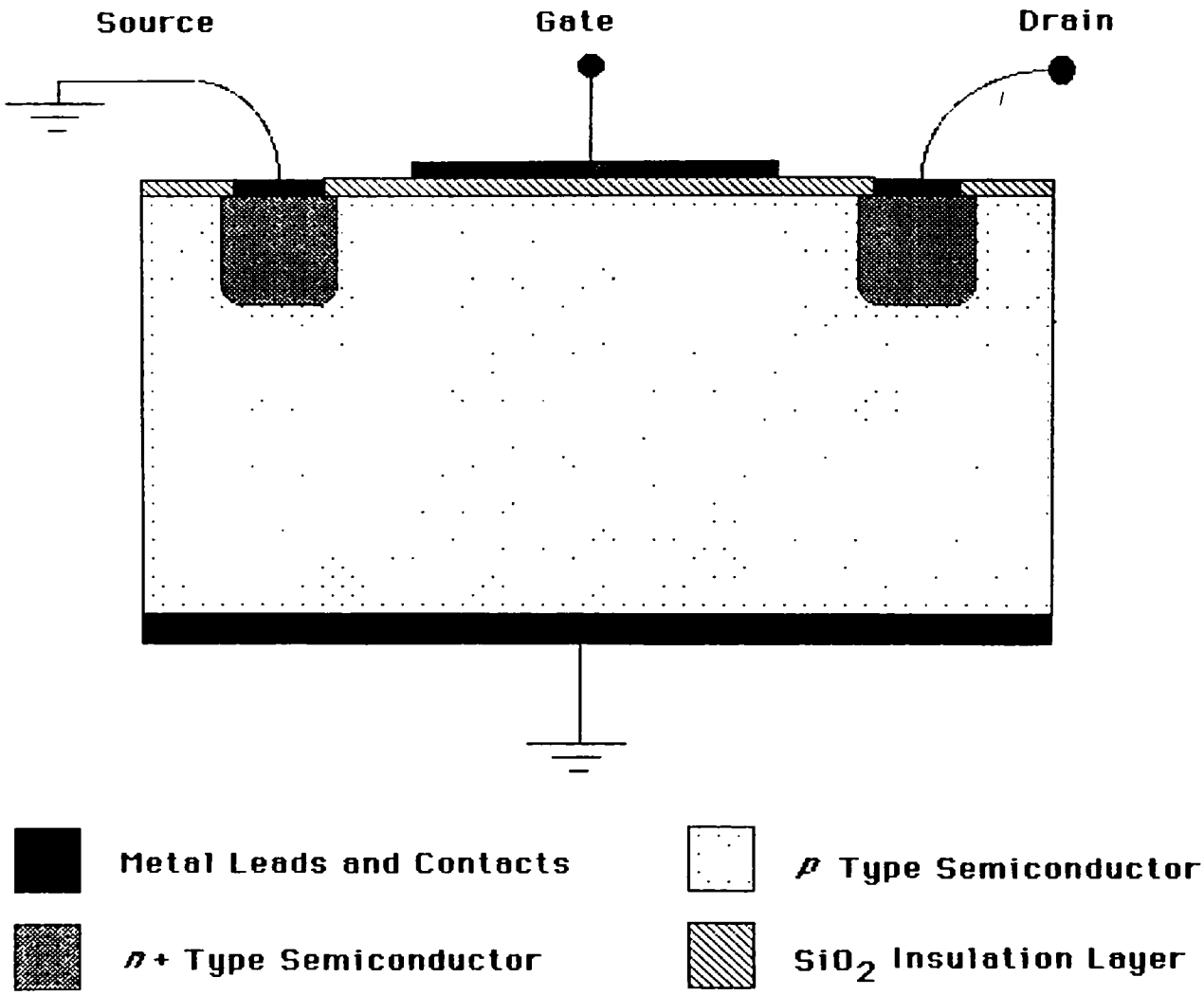


Figure 5: Cross-sectional view of an n-channel MOSFET.

to flow between the source and drain. The conductance of the inversion layer can be varied by varying the voltage applied to the gate.

In order for a current to flow in the inversion layer, there must be an additional potential applied across the source and drain (V_D). When this is the case, the conduction channel acts like a resistor whose resistance is inversely related to the electron charge density in the channel as long as the drain voltage is small compared to the gate voltage. However, as the drain voltage approaches the gate voltage, the average potential drop in this region, near the corner of the drain and gate contacts, will decrease. This also causes the local charge density to drop which will in turn increase the resistance of the channel. At some drain voltage, the potential in the channel near the drain will become small enough that the inversion layer can not be maintained and this region will no longer have an excess of electrons, but will simply be depleted of any charge carriers. In other words, there will be approximately equal numbers of electrons and holes. The drain voltage at which the event occurs is called the saturation voltage (V_{Dsat}). If V_D is further increased, the point at which the conduction channel meets the depletion region will gradually move toward the source. This process severely effects the current-voltage characteristics of the MOSFET. In the region where V_D is small, the current versus voltage curve is quite linear indicating a fairly constant conduction channel resistance. However, as the voltage approaches V_{Dsat}, the curve begins to bend over and flatten out. (see Fig. 6)[12] As the drain voltage passes saturation, the curve becomes almost horizontal indicating almost no more current gain with increasing V_D. This imposes an upper limit to the current (I_{Dsat}) that can pass through the channel for a given gate voltage . However, if the gate voltage is

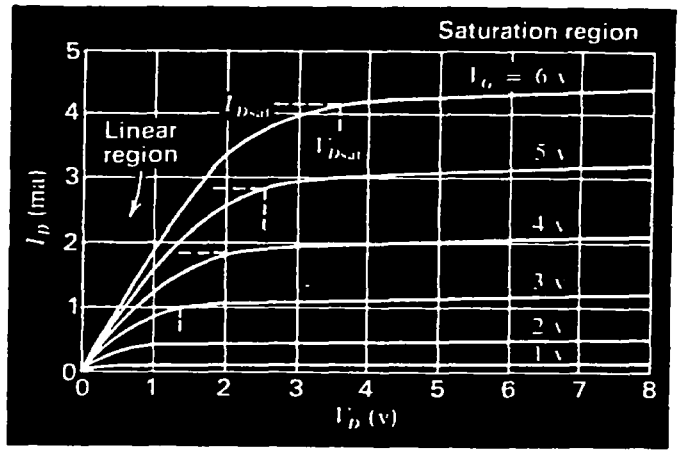


Figure 6: Typical current-voltage characteristics for an n-channel MOSFET.

increased then so is V_{Dsat} since that is only dependent on the voltage difference between the gate and drain. So, in this way, one can increase the maximum amount of current that can pass through the conduction channel.

In order for the MOSFET to function properly, a good insulation layer between the gate and semiconductor surface must be reliably grown. Silicon dioxide is an excellent choice for this job since it be easily grown on the semiconductor surface through thermal oxidation and has very good insulating properties. Before the idea for using SiO_2 as the insulating material was presented by Atalla's team at Bell labs (*see Historical Background section*), the limiting factor of the MOSFET was it's gate leakage current. Too many electrons would migrate through the insulator and leave the MOSFET through the gate instead of the drain. Using silicon dioxide however reduced the gate leakage current to levels low enough for the device to operate properly. A typical gate leakage current in a MOSFET is about 10^{-14} amps.[13] However, if the semiconductor surface is not prepared carefully and the insulating layer is not properly grown, the leakage current can be greatly affected which is why there has been so many studies concerning the surfaces of semiconductors.[14] (For a more detailed introduction to semiconductor surfaces, see A.S. Grove, *Physics and Technology of Semiconductor Devices* , chapters 9 and 10, John Wiley and Sons, 1967) It is the goal of the work presented here to provide further insight as to how silicon oxidation occurs in the initial stages. This will perhaps allow thinner oxidation layers to be reliably grown and further decrease the size of the MOSFET.

1-3 Overview of the Research

The oxidation of silicon is an integral part of semiconductor device production in many areas of microelectronics. Perhaps the most important of these is the production of MOS (metal-oxide-silicon) transistors where thermally grown SiO_2 is the gate dielectric material. (see above sections) There is a continued interest in making these transistors smaller, and therefore, thinner oxidation layers on the silicon substrate must be reliably grown. As a result, many studies in the last ten years have explored the silicon oxidation kinetics in the thin film range (20-500Å).[28,29,31,32] One factor that degrades the quality of the SiO_2 insulation layer is the formation of so called "native oxides" on the Si surface due to normal atmospheric exposure. However, it has been well established for some time that hydrogen terminates the dangling bonds on the silicon surface and thereby greatly reduces the native oxide growth and provides general protection against chemical attack.[15]

The simplest method of producing an H terminated Si surface is to expose the surface to aqueous HF acid.[23] This process both removes the native oxides and terminates the silicon surface bonds with H in mono-, di-, and trihydride phases depending on the crystal face exposed to the solution. Since H terminates the (111) plane with monohydride phase hydrogen, and the di-hydride phase only occurs at crystal corners or defects and the trihydride phase only occurs as an adatom; complete monohydride phase termination indicates an atomically smooth surface.(see Fig. 7) One practical problem with using HF treatment for TDS experiments, however, is that there is a great deal of background hydrogen in the atmosphere. Even ultra-high vacuum chambers have a relatively high H_2 background. Therefore, in

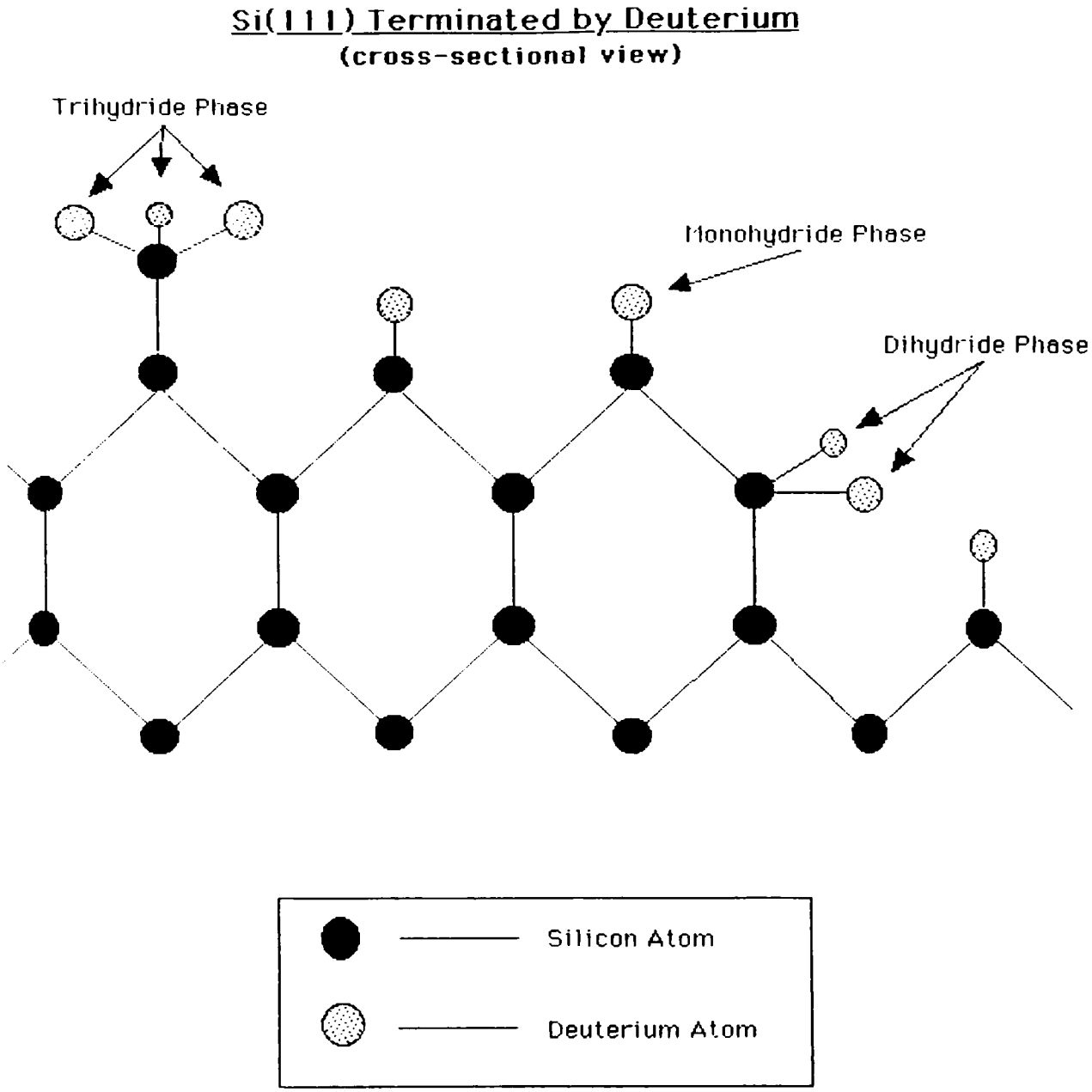


Figure 7: Cross-sectional view of a deuterium terminated Si(111) surface. As can be seen, the bonds at the step sites are terminated with the dihydride phase, while the flat (111) surface is terminated with the monohydride phase. The trihydride phase only occurs when a single Si atom sticks up above the surface exposing it's other three bonds.

order to make more accurate observations, we terminated the silicon surface not with hydrogen, but deuterium, since it has the same chemistry, but a far lower background level.

It has recently been reported that boiling H terminated Si(111) in water leaves the sample with only monohydride phase H termination, and therefore a smooth surface.[16] We have used this technique substituting heavy water (D_2O) for the normal water and found that this did leave the surface smooth and D terminated, but also left a rather significant oxidation layer. This however was eliminated by adding a very small amount of DF to the boiling D_2O.

After an effective method of sample preparation was found, we exposed the samples to an oxidizing environment (clean room air) at various temperatures. We found that the SiO layer growth has a surprisingly linear dependence on oxidizing temperature for both smooth and rough surfaces within the temperature range of 125 to 325°C. During this same temperature range, the D_2 peak intensity decays in a roughly exponential fashion. These results, as well as others reported here, seem to support an initial oxidation model in which O_2 first inserts into the Si backbonds between the first and second atomic layers of the Si crystal.

Here I report the results of a study investigating the dependence of initial oxide growth kinetics on D terminated Si(111) as a function of the temperature of the oxidizing environment with an oxidation layer thickness of 0-15Å. Our general research plan included the completion of the following tasks:

1. Determine the conditions required in order to create an atomically smooth, 100% D terminated, and oxide free Si(111) surface.

2. Study the oxidation reactions on the smooth and rough D terminated surfaces.

A smooth crystal surface is defined to be one which is composed almost entirely of flat planes. A rough surface contains many step sights or transitions from one crystal plane to the one above or below. (See Chap. 2: Experimental Procedures for a discussion on how these surfaces are created.)

The method chosen for these studies is Thermal Desorption Spectroscopy (TDS). In this type of detection method, the species attached to the surface of the sample are desorbed thermally, ionized, and injected into a quadrupole mass spectrometer via an electric field. (see Chap. 2: Experimental Procedures and Appendix) TDS has several advantages in this type of study in that the hydride, deuteride, and oxide species can be observed simultaneously. Also, the monohydride and dihydride phases of H(D) can be resolved.[25]

Chapter 2: Experimental Procedures

2-1 Experimental Equipment

All of the experiments for this thesis were performed in Tsukuba, Japan at the Nippon Electric Corporation (NEC) Laboratory and the Photon Factory (a synchrotron radiation ring) of the High Energy Physics Laboratory which is a government supported facility. While the electromagnetic radiation produced by the PF ring was not used in these experiments, the ultra high vacuum chambers there were needed to conduct the TDS experiments. All of the chemical preparation and baking of the samples was done at the NEC lab in one of the clean rooms. Silicon surfaces are very sensitive to atmospheric conditions which is why sample preparation must be done in clean rooms. They are laboratories which are sealed to the outside environment and with the air being constantly filtered to reduce the level of dust to a minimum. All workers in the clean rooms wear dust suits which cover all parts of the body except the face and hands. Usually eight samples were prepared each day and packed in an inert N_2 environment for transport to the vacuum chamber and TDS apparatus at the Photon Factory (PF). After arriving at the PF ring, 3 or 4 samples were loaded onto holders and placed in the introduction vacuum chamber which was sealed and evacuated by a turbomolecular pump. Meanwhile, the reaction chamber and quadrupole mass spectrometer shroud were cooled with liquid nitrogen in order to obtain lower base pressures. The ultra-high vacuum (UHV) consists of 3 chambers of decreasing base pressures.(see Fig. 8) The reaction chamber has a base pressure of about 8×10^{-10} Torr. The desorbed species were

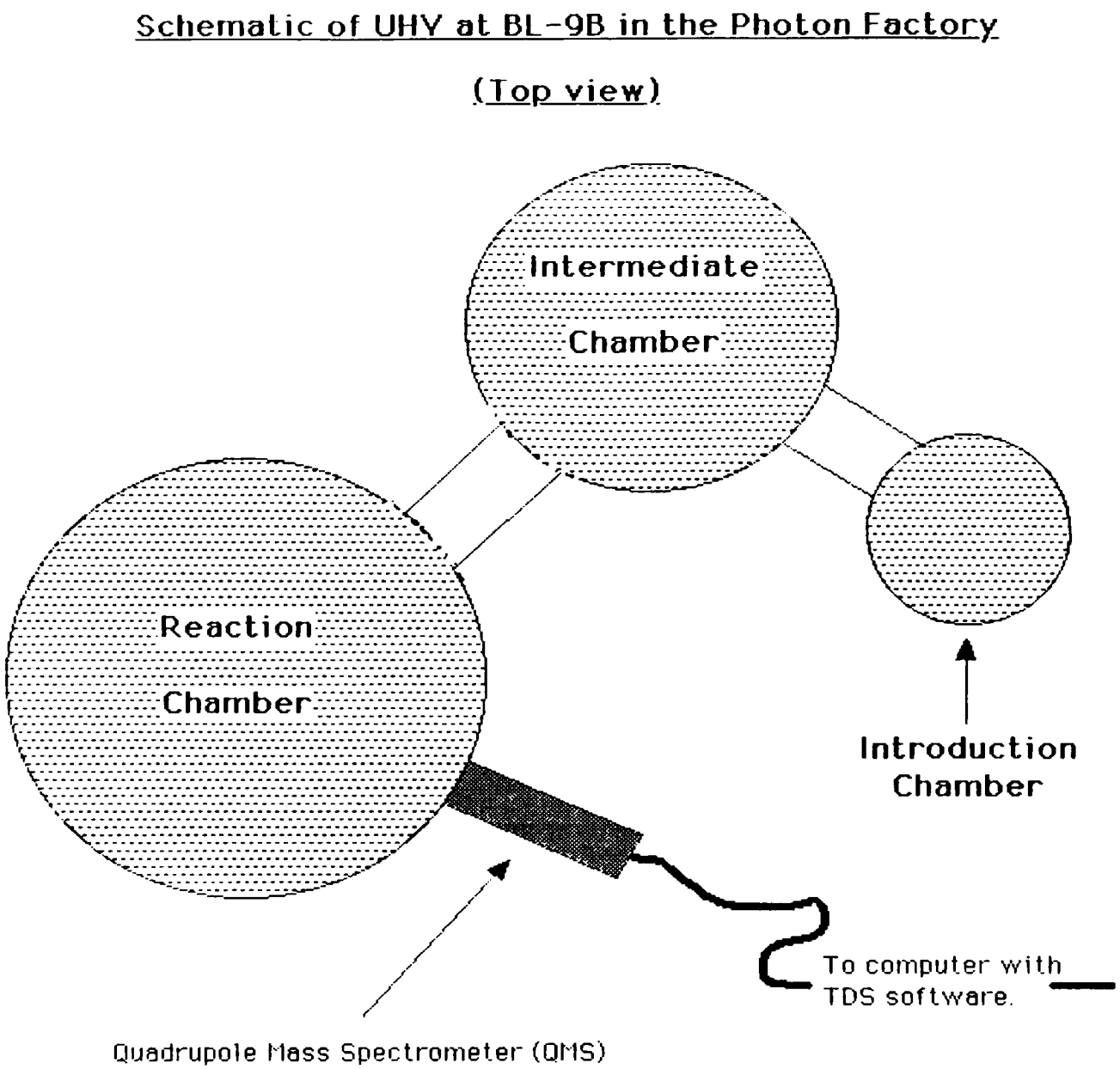


Figure 8: Schematic (top view) of the ultra high vacuum chamber at the Photon Factory in Tsukuba, Japan.

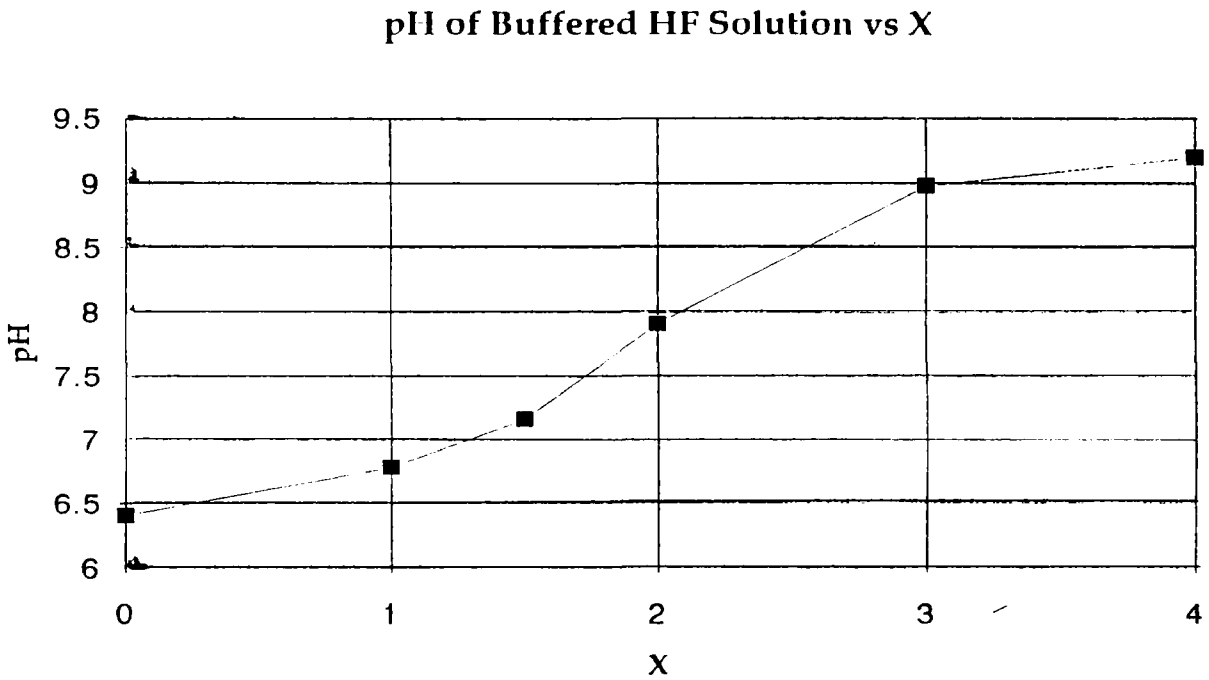


Figure 9: Plot of the pH of the buffered HF solution as a function of X, where X is the ratio of NH_4OH added to the solution.

detected by a quadrupole mass spectrometer (QMS - see Appendix) which had a base pressure of 3×10^{-10} Torr; and samples were heated using an electrical resistance heater.

The silicon samples were loaded onto the holders and placed in the introduction chamber which was sealed and evacuated by a turbomolecular pump until the pressure reached about 9×10^{-6} Torr. Each sample was then individually transferred to an intermediate chamber and then to the reaction chamber where it was positioned approximately 2 cm in front of the orifice of the QMS shroud. Pressure in the reaction chamber and QMS shroud were typically 1×10^{-9} and 7×10^{-10} Torr respectively. The samples were heated at a rate of 50°C per minute with TDS data recorded from 200°C to 1100°C. Nine mass numbers were observed: m/e = 2, 3, 4, 18, 19, 20, 44, 45, 46 which correspond to H_2, HD, D_2, H_2O, HDO, D_2O (or HF), and three isotopes of SiO, respectively.

2-2 Sample Preparation

The 30x30 mm^2 samples were cut from 4" lightly B doped Si(111) wafers cut with a 3 degree offset. These samples were chemically cleaned by boiling in a solution of $NH_4OH : H_2O_2 : H_2O$ = 1:1:5 for 15 minutes (APM treatment) and rinsing in deionized water for 15 minutes. They were finally packed in a nitrogen purged environment. Daily preparation included a one minute dip in dilute hydrofluoric acid (dHF) (1.6% in water)to remove any native oxide surface layer and hydrogen terminate the dangling bonds of the Si(111) surface. For investigating the H-D exchange rate, the samples were then dipped in a dilute DF (dDF) solution (1.6% in heavy water) for varying lengths of time.

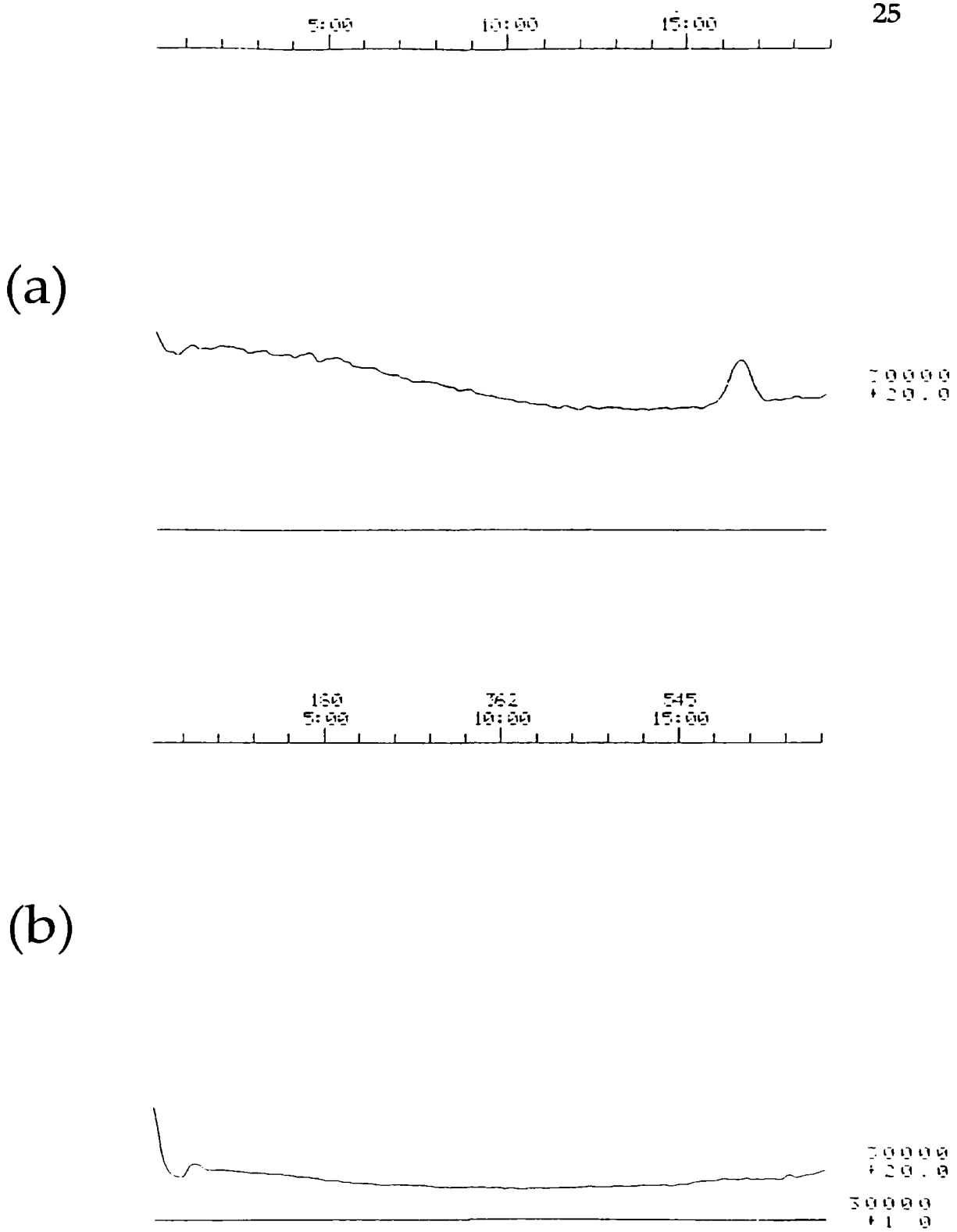


Figure 10: (A) SiO TDS of a sample treated by D_2O boiling, (B) SiO TDS of a sample treated by H_2O boiling.

2-3 Buffered HF Treatments

For investigating the pH dependence on surface smoothing, dHF treated samples were given a 5 minute dip in a buffered pH solution of varying pH. The buffered HF solutions consisted of NH_4F : HF : H_2O : {pH varying agent} = 7:1:(7-x):x where x is adjusted to change the pH. For x=0, the pH=6.3 ; NH_4OH was added to increase the pH, and HCl was added to lower the pH. A plot of pH dependence on x when the pH varying agent is NH_4OH is shown in Fig. 9. Samples were again packed in a nitrogen purged environment for transportation to the PF ring.

2-4 Heavy Water Boiling Experiments

Daily treatment of the samples for the heavy water boiling experiments was 10 minutes in dDF (1.6% in heavy water) followed by boiling in D_2O (99.75%) for various lengths of time from 0 to 30 minutes. Once conditions were found which left the Si surface both smooth and 100% D terminated, the roughening rate of dHF (dDF) was studied by exposing the smooth samples to these solutions for various lengths of time. Also, in order to study the removal of the di- and tri-hydride phase deuterium during boiling D_2O treatment, the rough D terminated samples were again exposed to boiling heavy water for varying lengths of time.

2-5 Preparation of Atomically Smooth and Oxide Free Surfaces

An important result reported above is the success in making an atomically smooth and 100% D terminated Si(111) surface. We followed the general procedure reported by Watanabe et. al., which used a dHF dip

followed by a boiling water treatment to create a smooth H terminated surface.[16] Since we wanted a D terminated surface, dDF and heavy water were used instead. Watanabe et. al. checked for surface oxidation by 2p photoelectron spectroscopy and found none. It was assumed that substituting deuterium for hydrogen would have no effect on any oxidation effects. However, during the course of this study, we paid close attention to the SiO species and discovered that our samples treated by dDF -> boiling D_2O did in fact have a rather significant oxidation layer of 10-15Å. Our procedure was checked by attempting to reproduce the results of Watanabe et. al. in making a H terminated surface and no SiO peak was found for the H terminated samples. Figure 10 shows the SiO TDS spectra of both H and D terminated samples prepared by boiling treatments.

The reason for this residual oxidation layer left by heavy water boiling has not yet been determined. The chemistry of H and D is the same, with the only difference between the two being their masses. However, since our present study is of the oxidation effects of D terminated Si surfaces, it was paramount that our samples had no initial oxidation layer. Therefore, we concentrated only on preventing the formation of the layer, not the reason for it's existence. We found that boiling the samples in very dilute DF instead of pure heavy water produced a reasonably good balance between smoothing the surface and preventing the formation of the oxide layer. For creating a smooth and oxide free surface, the procedure was: 10 minutes in dDF, 10 minutes in boiling dDF of varying concentrations from 0.16% to 0.00016%. Figure 11 shows the SiO and D_2 TDS spectra of samples boiled in varying concentrations of very dilute DF. A concentration of 0.005% of DF in heavy water produced a surface that was 95% smooth and oxide free.

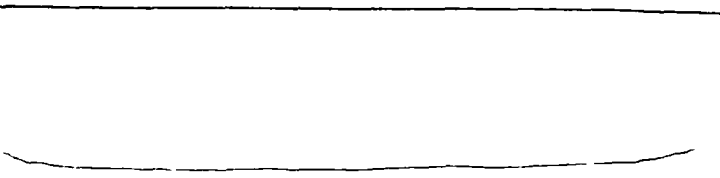
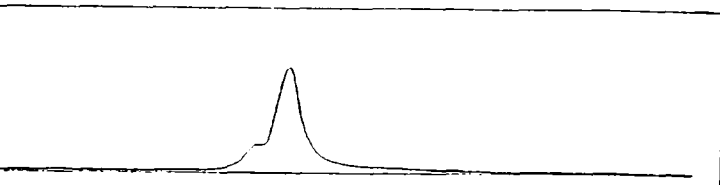
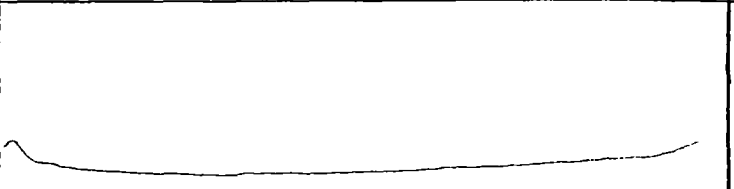
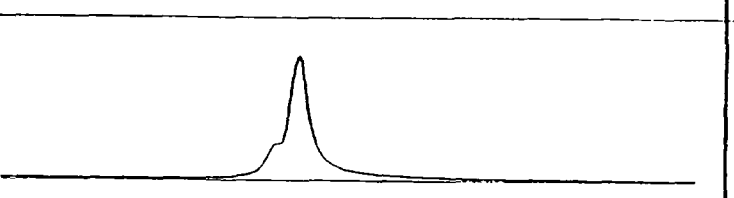
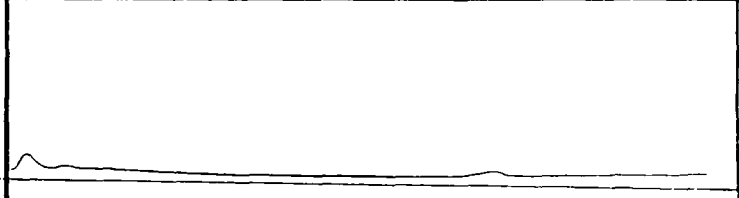
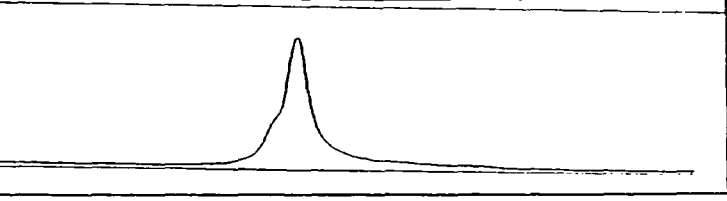
Boiling DF Concentration	Species	TDS Spectra	m/z
0.008%	SiO		44
	D_2		4
0.005%	SiO		44
	D_2		4
0.0016%	SiO		44
	D_2		4

Figure 11: SiO and D_2 TDS spectra of samples prepared by dDF (1.6%) for 10 minutes -> then boiled in very dilute DF of varying concentrations as indicated. It was decided that 0.005% produced the best compromise between a smooth *and* oxide free surface.

Subsequently, data in the temperature dependence experiments labeled smooth were prepared by 10 minutes in dDF (1.6%) followed by boiling in very dilute DF (0.005%) for 10 minutes.

2-6 Oxidation Experiments

Daily treatment of the samples for the initial oxidation experiments was as follows: for rough surface samples -> 10 minute dip in dDF (1.6%), 10 minute boil in D_2O (99.75%), 1 minute dip in dDF (1.6%), finally heat for one hour at various temperatures. For smooth surface samples -> 10 minute dip in dDF, 10 minute boil in very dilute DF (0.005%), and heat for 1 hour at various temperatures. The samples were heated in clean ovens with a temperature range of 20 to 350° C. The oxygen partial pressure in the ovens is 0.21 atms.

Chapter 3: Results and Discussion

3-1 General Features of TDS Spectra

Typical TDS spectra for dDF treated samples exhibited three peaks for m/e = 2, 3 (H_2, HD) and two peaks for m/e = 4 (D_2) as shown in Fig. 12. It should be noted here that while all of the TDS spectra shown in this thesis are raw data, they exhibit typical and reproducible features. The low temperature peak (T_S=510 to 550°C) was ascribed to the dihydride and trihydride phases while the next peak (T_S=600 to 610°C) was ascribed to the monohydride phase.[17] The third peak in the H_2 and HD channels (T_S=775 to 825°C) has not yet been accounted for and could possibly be the subject of future study. Because the Si(111) surface has a single dangling bond, a perfectly flat H terminated Si(111) surface will produce only the monohydride phase peak and no dihydride peak will appear. Therefore, a measure of the surface smoothness was obtained by taking a ratio of the low temperature (dihydride) peak intensity to the higher temperature (monohydride) peak intensity.

3-2 Buffered HF pH Dependence on the Surface Smoothness

One of the methods we tried in making an atomically flat, D terminated surface was the use of buffered and pH modified HF treatment, followed by dDF treatment. The buffered and pH modified HF treatment has been shown to leave the H terminated surface flat.[22] Samples treated with dilute HF and buffered HF only showed TDS structures in the channels

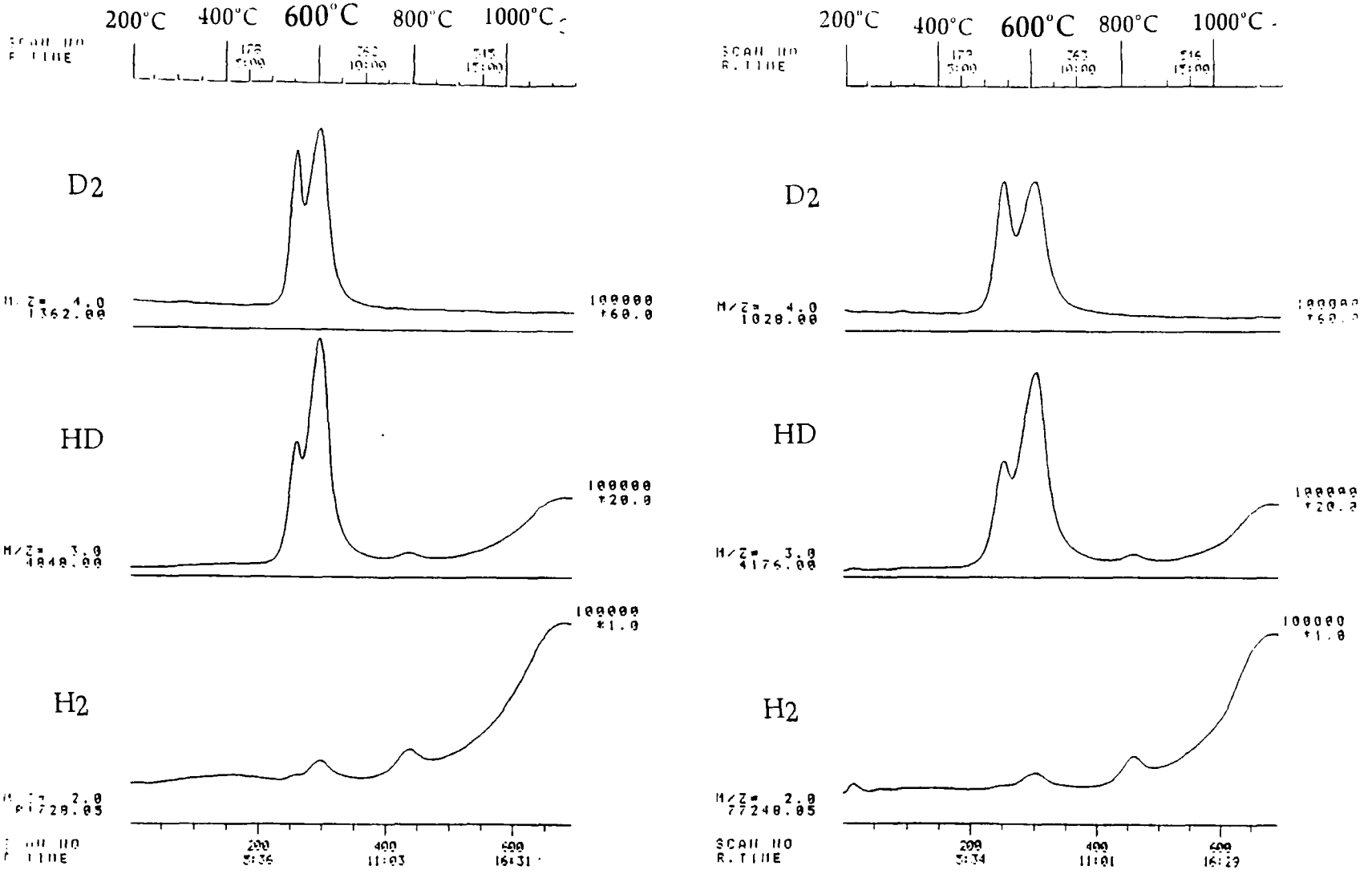


Figure 12: Typical TDS spectra of samples treated by (A) dHF (1 minute) -> dDF (30 minutes), and (B) dHF (1 minute) -> dDF (60 minutes).

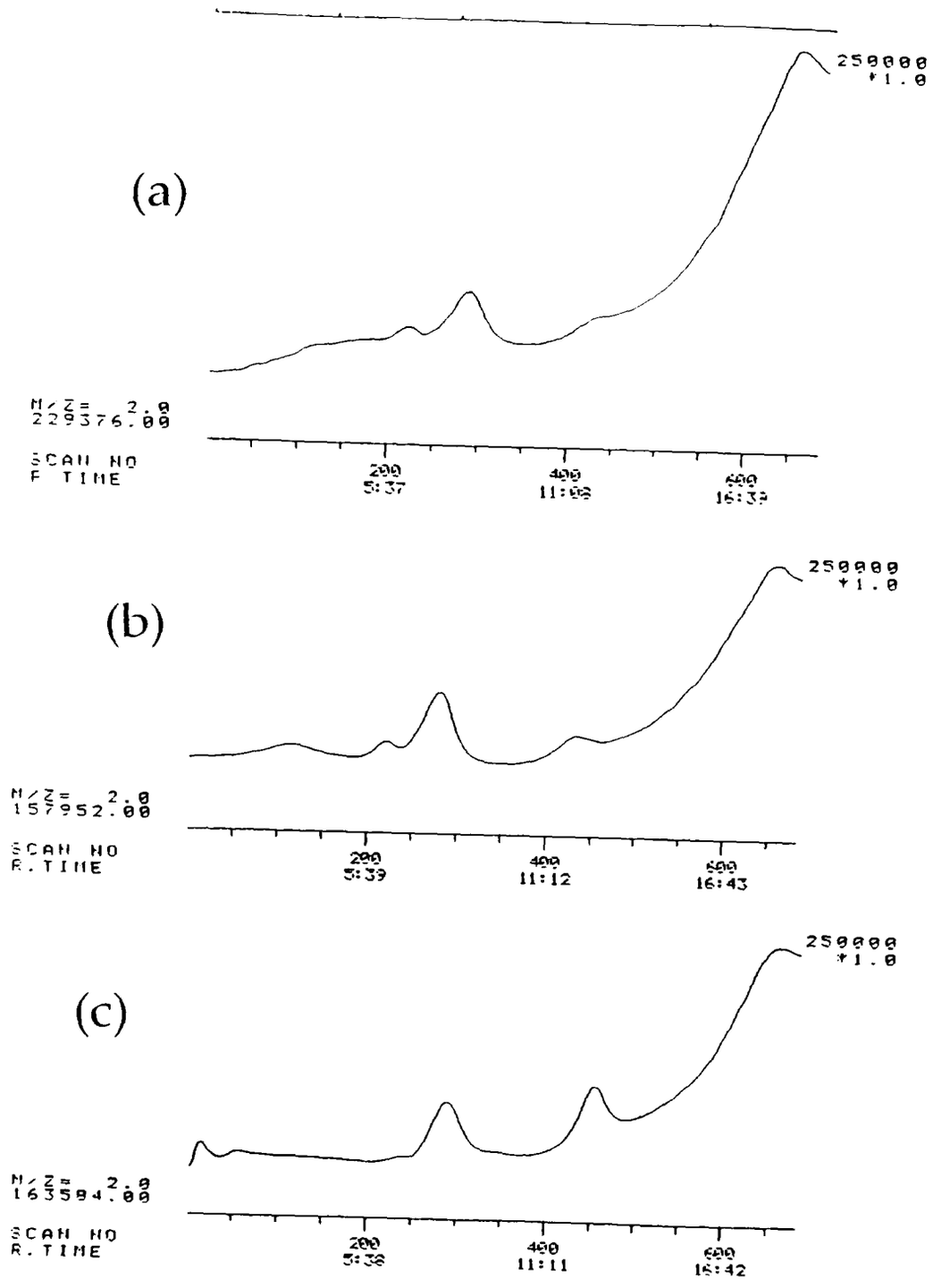


Figure 13: H_2 TDS spectra for Si(111) samples prepared by (a) pure dHF only (pH=3.2), (b) buffered HF (pH=6.4), and (c) buffered HF (pH=9.0).

corresponding to m/e = 2 and 3, that is H_2 and HD. The same three H_2 peaks were observed for all cases. As stated above, a measure of the roughness of the Si(111) surface is the ratio of the intensity of the di-/tri-hydride to monohydride peaks. Figure 13 shows several TDS for varying pH. It is clear that the intensity of the dihydride peak decreases with increasing pH, which is shown graphically in Fig. 14. Once the baseline is chosen, the intensity ratio can be measured by the peak heights within the limits of uncertainty. Our results agree with previous studies [18] that show a buffered HF solution of pH>7 produces an atomically flat Si(111) surface. However, we wanted to terminate the surface with deuterium, and since the dDF treatment was found to reroughen the sample, we decided this method was not appropriate for our experiments.

3-3 H/D Exchange Reaction

As Fig. 12 shows, the spectra for 30 minutes and 60 minutes exposure to dDF after dHF treatment are very similar. The intensity ratio of H_2 to D_2 monohydride phase peaks changes by only approximately 10% from t=30 to 60 minutes. It is therefore concluded that the H-D exchange reaction is very close to saturation after 30 minutes and future samples can have dDF exposure times of less than this time and still accurately plot the deuterium coverage as a function of time. It should also be noted here that the H-D exchange was not complete even after one hour exposure. This will be discussed in further detail below.

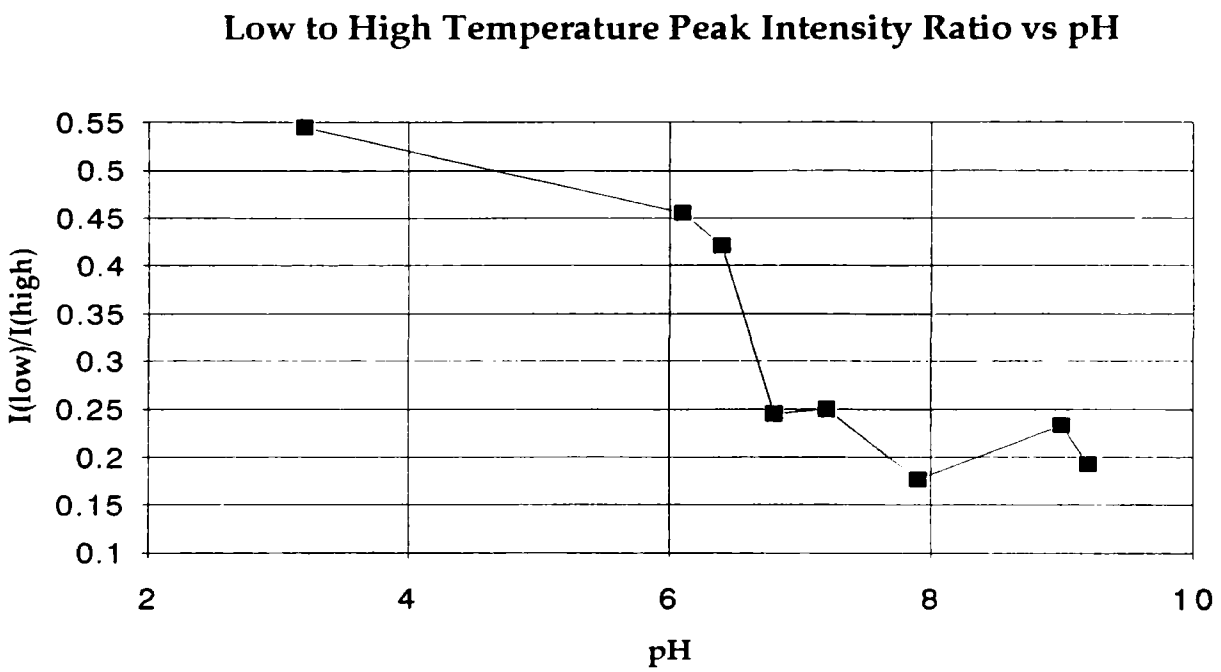


Figure 14: Plot of the dihydride to monohydride peak intensity ratio as a function of pH.

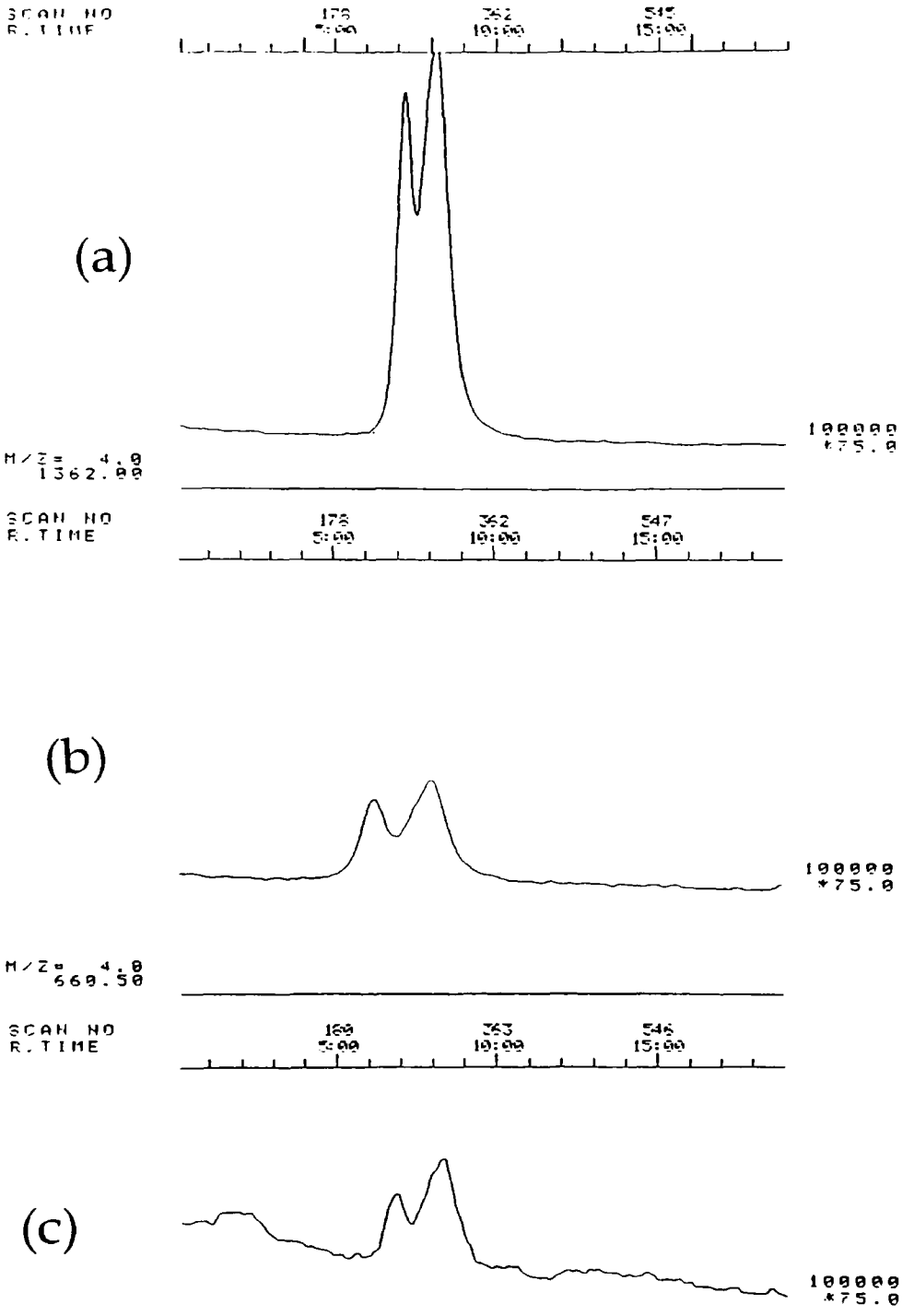


Figure 15: D_2 TDS spectra for Si(111) samples prepared by (a) pure dHF -> dDF (30 minutes), (b) buffered HF (pH=7.8) -> dDF (30 minutes), (c) buffered HF (pH=7.8) -> dDF (10 minutes).

3-4 Roughening of the Smooth Si(111) Surface

An interesting feature for D_2 should be noted here; the dihydride peak increases in intensity relative to the monohydride peak as the dDF exposure time increases from 30 to 60 minutes. One possible explanation for this is the H-D exchange rate is higher for the dihydride phase, which only forms at the crystal corners (see Fig.s 6 & 16), than the monohydride phase.[19] Further evidence for this possibility arises when the TDS spectra for pure dHF -> dDF treatment (Fig. 15a) and buffered HF (pH=7.8) -> dDF treatment (Fig. 15b) are compared. The intensity levels for the D_2 peaks are considerably lower for the sample treated with the buffered HF, although the dihydride over monohydride peak intensity ratio is almost the same. One possible explanation for this result is that the relatively flat Si(111) created by the buffered HF solution is reroughened by the dDF treatment. More precisely, the number of dihydride sites available at the beginning of the dDF treatment is very low, and since the monohydride H-D exchange rate is slower than the dihydride, the whole H-D exchange rate is slower for the buffered HF treated sample resulting in smaller intensity peaks for the same dDF exposure time. However, since the chemistry of HF and DF is almost the same, the surface is made rough again during the dDF exposure, which creates more dihydride phase sites. This results in a dihydride peak of similar relative intensity as that for the non-buffered HF treated sample.

This hypothesis is strengthened when the TDS spectrum for a buffered HF (pH=7.8) -> dDF treated sample with a shorter exposure time is observed (Fig. 15c). As expected, the di- to monohydride peak ratio is smaller by about 15% with the shorter exposure time. What is interesting however is that the monohydride peak intensity actually decreased slightly with the

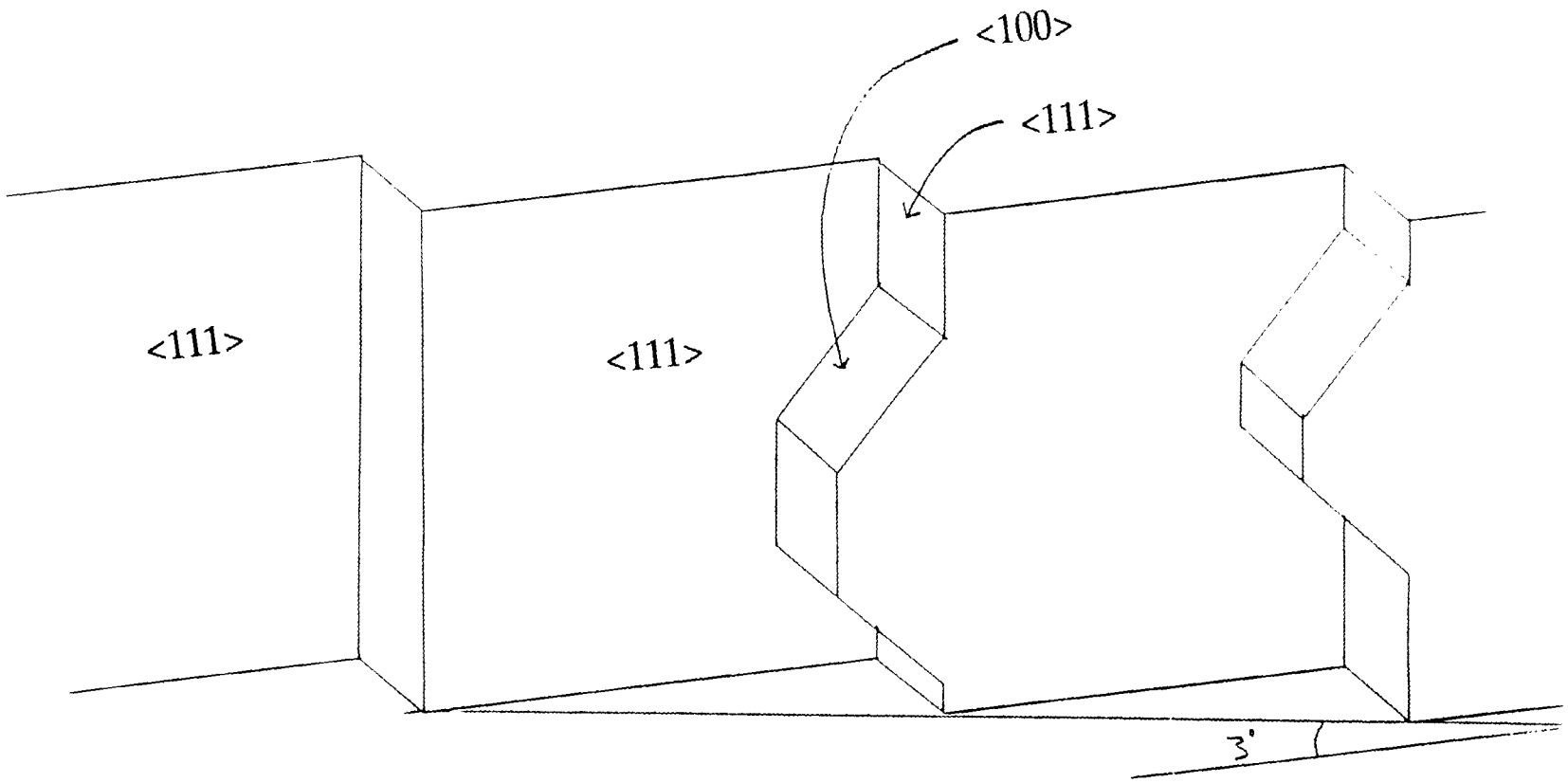


Figure 16: Example of a rough Si(111) surface cut with a 3° offset.

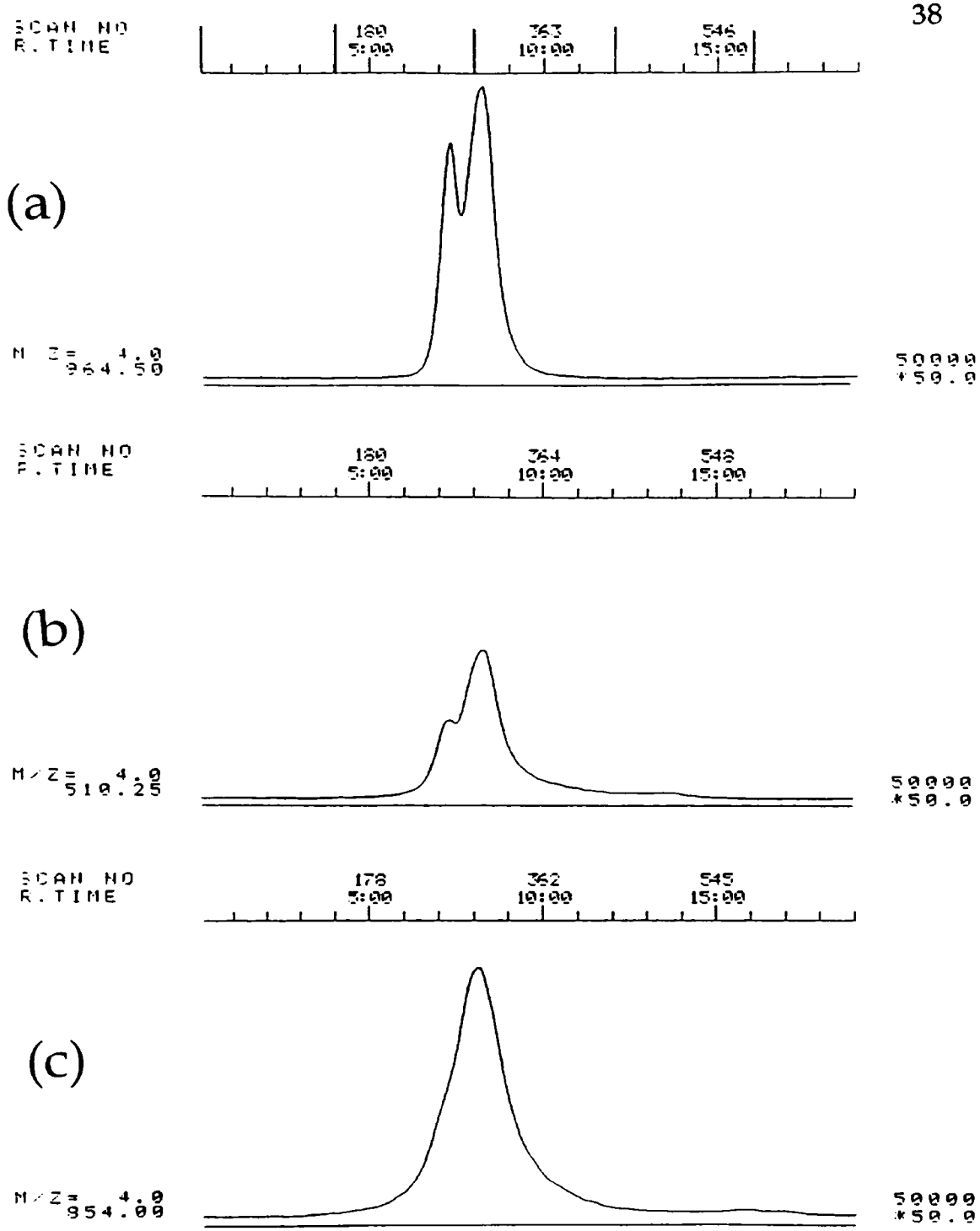


Figure 17: D_2 TDS spectra of Si(111) prepared by dDF for 10 minutes -> boiling in D_2O for t seconds where (A) t=0 , (B) t=30, (C) t=600.

longer dDF exposure time, while the dihydride peak intensity increased. Therefore, it seems that the number of the monohydride phase sites on a flat Si(111) surface created by the buffered HF solution is decreased as the dDF exposure increases. Meanwhile, the reroughening process increases the number of dihydride sites (crystal corners in the (100) plane). A close examination of the Si(111) crystal lattice reveals that a corner created in one plane, the (100) plane, results in a dihydride site, while a corner created in the 135° plane, the (11$\underline{1}$) plane results in a monohydride site. (Fig. 16) So far, we can not determine unequivocally that there is any preference of direction (with regard to crystal orientation) in the roughening process. From the limited data already taken, the slight decrease in monohydride peak intensity with increasing dDF exposure time would seem to indicate that the roughening process is a little faster in the (100) direction than the (11$\underline{1}$). Comparing this data to similar experiments using Si(100) might help shed some light on this question.

3-5 Effects of Boiling D_2O

The effect of boiling the Si samples in heavy water was two fold: first, the rough surface created by the dDF treatment was smoothed; and second, the deuterium termination of the surface dangling bonds was completed. As Figs.17 and 18 show, the TDS spectra of increasing boiling length show a decrease in intensity of both the D_2 dihydride (low temperature) peak and H_2 peaks. The fact that hydrogen was observed without HF treatment means that either it was adsorbed on the surface during APM treatment and not removed during dDF exposure, or it was adsorbed during the dDF treatment itself. For either case, a significant H contamination of the dDF solution is

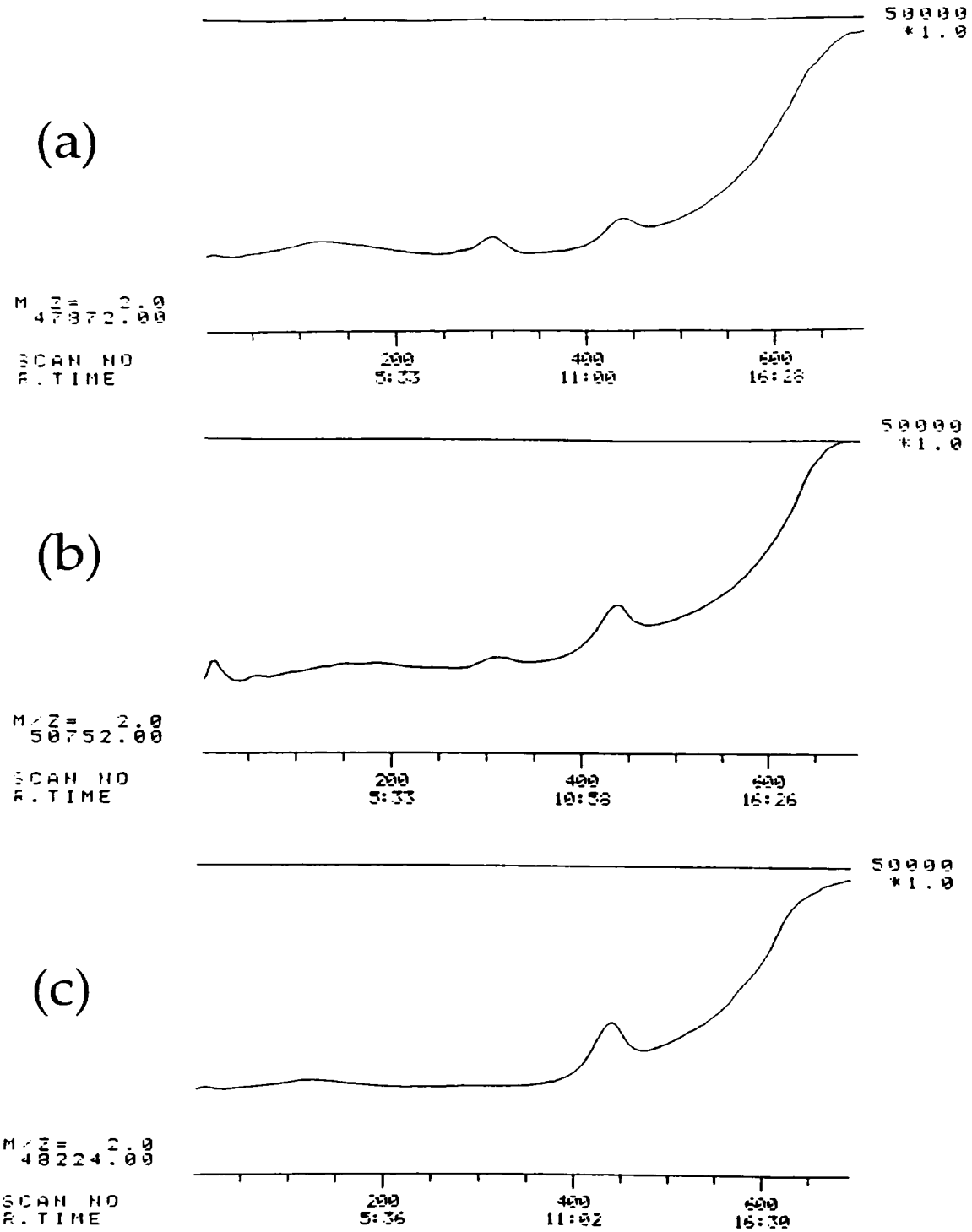


Figure 18: H_2 TDS spectra of Si(111) prepared by dDF for 10 minutes -> boiling in D_2O for t seconds where (A) t=0 , (B) t=30, (C) t=600.

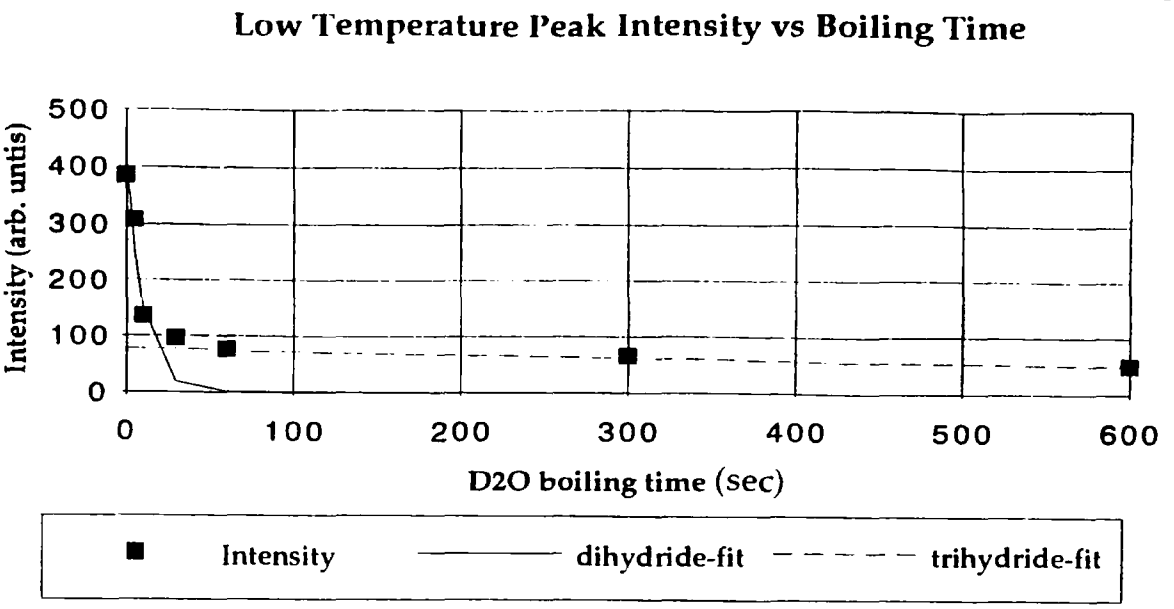


Figure 19: Intensity of the low temperature D_2 peak as a function of boiling D_2O exposure time. Intensity is measured as the height of the di/trihydride peak above the monohydride peak.

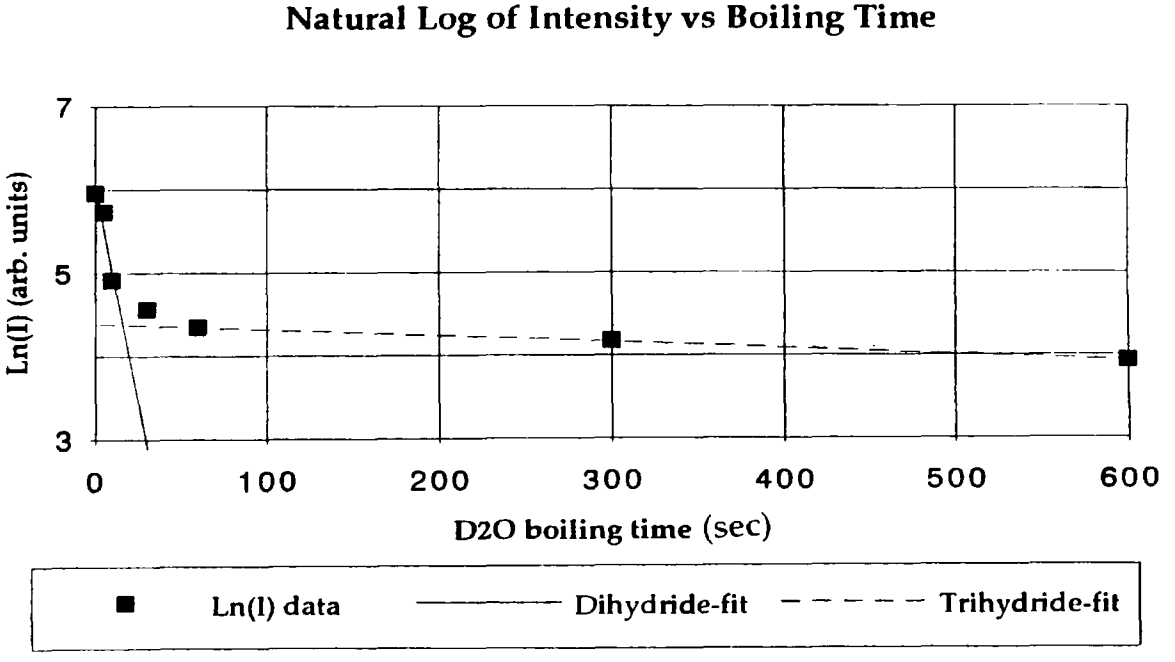


Figure 20: Natural log of the intensity of the di/trihydride peak as a function of boiling time.

very likely. This possibility is discussed further below. After a boiling time of 10 minutes, however, hydrogen had been removed and the spectra showed no further changes for either H_2 or D_2 species. It is therefore concluded that the surface smoothing rate and H/D exchange rate have reached saturation at this point. By saturation, it is clear from Fig. 18c that the H_2 mono- and dihydride peaks are below detectable levels and from Fig. 17c that the D_2 dihydride peak is also reduced to negligible levels.

3-6 Two Distinct Smoothing Rates

A plot of the intensity of the D_2 low temperature peak as a function of boiling D_2O dipping time is shown in Fig. 19. The intensity is measured as the height of the dihydride peak above the monohydride peak background which is extrapolated by assuming it is symmetric about the maximum. It is a reasonable assumption that the D_2 intensity decreases in an exponential fashion, so in order to determine the decay rate, the natural log of the intensity was plotted as a function of time (see Fig. 20). If there is a single decay rate, the natural log plot should result in a single line while two rates would produce two lines. From this plot it is clear that there are two distinct decay rates; one from 0 to about 30 seconds, and one from 30 to about 600 seconds. A linear least squares fit to the data in the log plot shows the first decay rate to be: 0.104 sec^{-1} and the second to be 7.61×10^{-4} sec^{-1}. The fact that two decay rates were observed agrees with previous reports that the low temperature TDS peak shows not only the dihydride but also the trihydride phase as well.[20] This report of Greenlief et al. compares studies of similarly prepared Si(111) samples by TDS and static secondary ion mass spectrometry (SSIMS) which can resolve mono-, di-, and trihydride phases. Their results

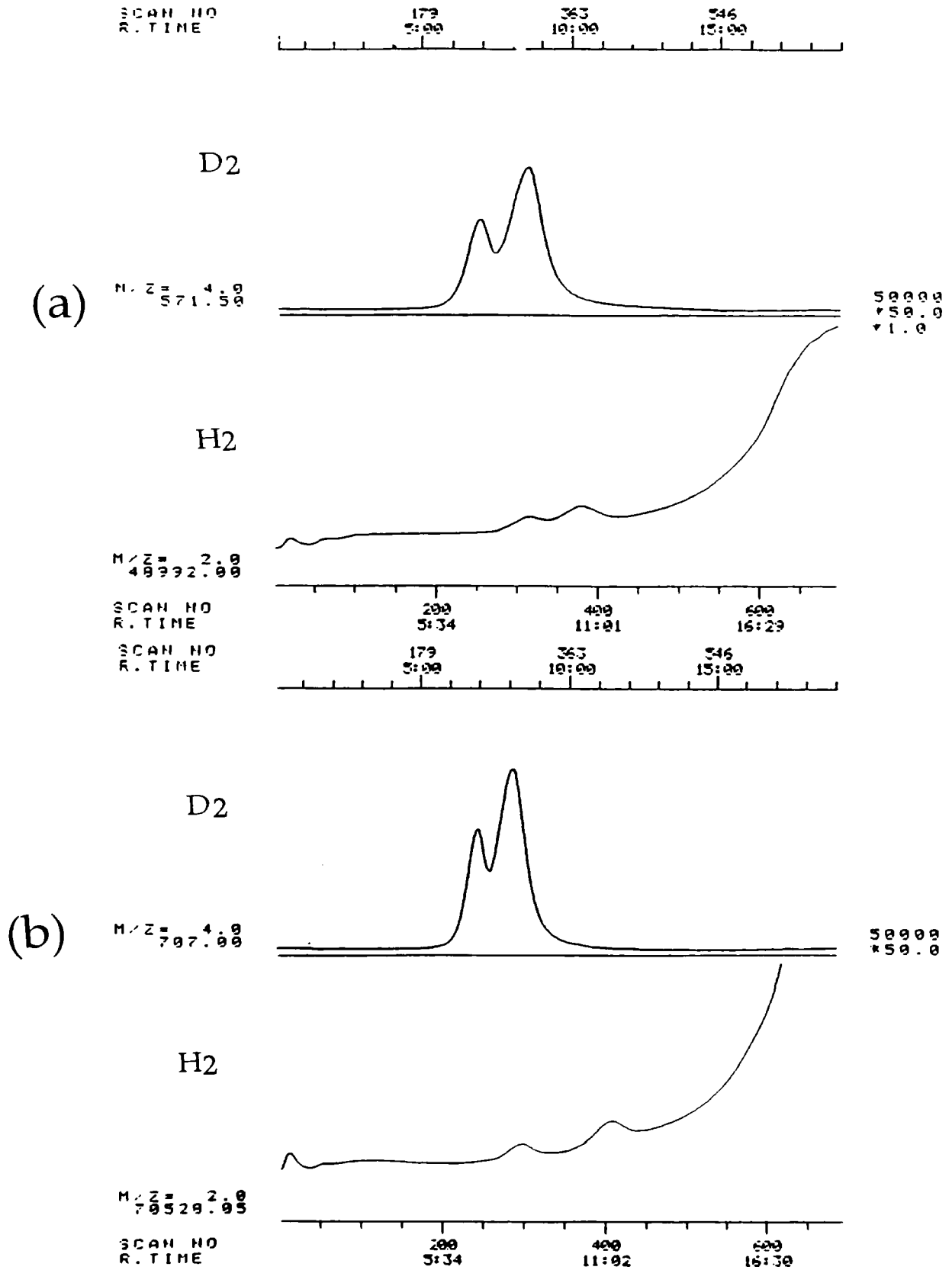


Figure 21: TDS spectra of both H_2 and D_2 prepared by exposing the smooth, D terminated Si sample to dDF for (A) 300 seconds and (B) 30 seconds.

showed that the low temperature TDS peak was created by both the di- and trihydride phases. In agreement with another study was the fact that the decay rate for the dihydride phase is much faster than that of the trihydride phase.[21] Comparison of the rates calculated above shows that the dihydride rate is about 140 times faster than the trihydride rate. Also in agreement with Watanabe et al., we observed that the dihydride phase D is almost completely removed after only 30 seconds in the boiling treatment, while the trihydride phase deuterium are not removed until about 10 minutes. It should be noted that Watanabe et al. used polarized infrared absorption spectroscopy which can also resolve the mono-, di-, and trihydride phases separately.

We therefore conclude that the decay rate between 0 to 40 seconds corresponds to the removal of the dihydride phase deuterium and the decay rate between 30 and 600 seconds corresponds to the removal of the trihydride deuterium. Calculations show that the dihydride rate (R_D) is 0.104 sec^{-1}, and the trihydride rate (R_T) is 7.61×10^{-4} sec^{-1}. Furthermore, these rates were used to calculate the initial intensities (0 seconds in boiling heavy water) of all three phases. The results given as percentages of the total initial intensity (I_o) are: monohydride = 66% I_o, dihydride = 27% I_o, and trihydride = 7% I_o. These intensities are for samples treated with dDF (1.6%) for 10 minutes and are of course dependent on treatment conditions.

3-7 Roughening effect of dDF

In order to find out whether the smooth, D terminated Si surface was affected by acidic conditions or not and also to determine if H contamination was significant, such samples were dipped in dDF for various time

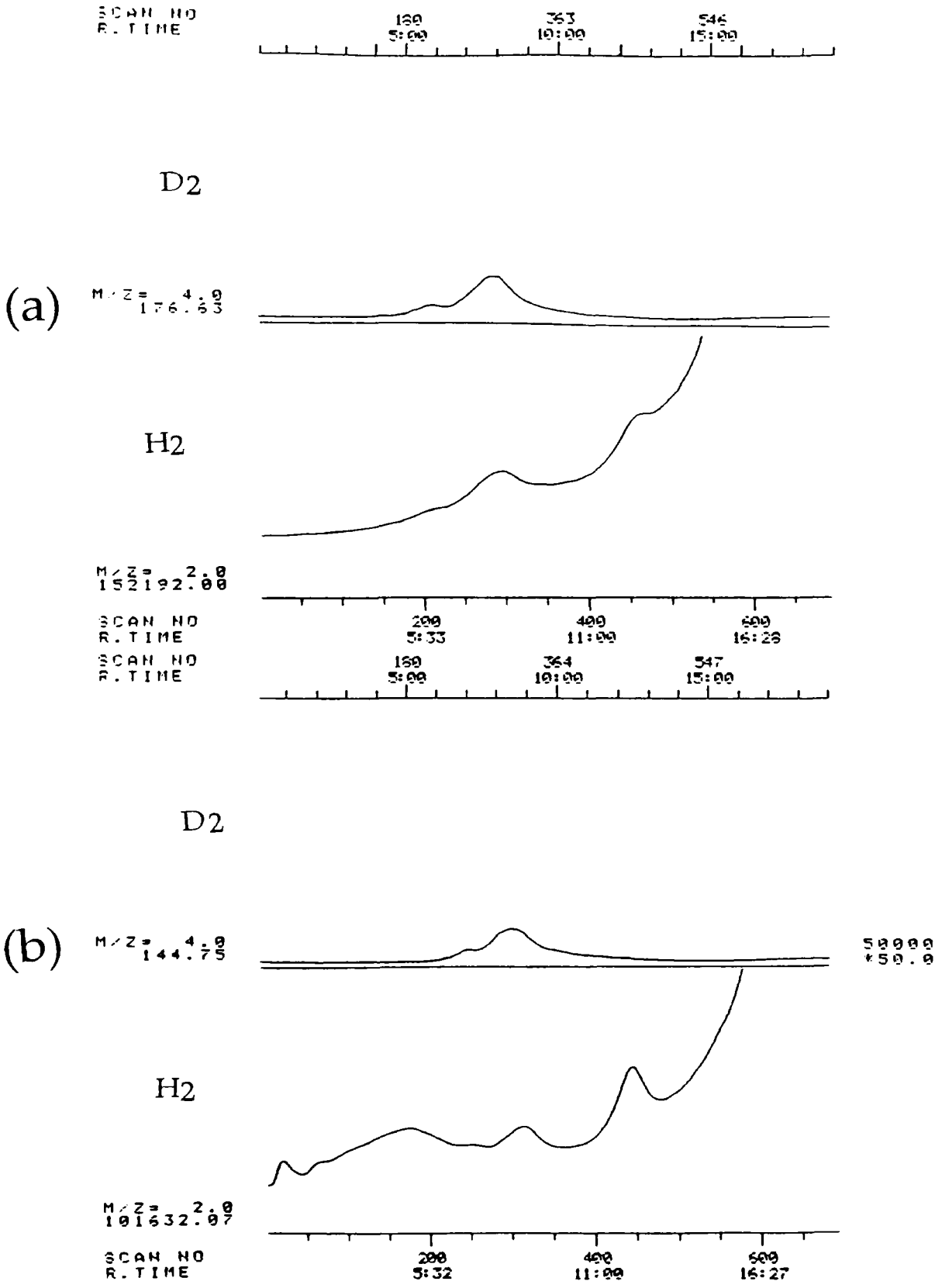


Figure 22: TDS spectra of both H_2 and D_2 prepared by exposing the smooth, D terminated Si sample to dHF for (A) 5 seconds and (B) 10 seconds.

durations. Figure 21 shows spectra for several different dDF exposure times. As can be seen, the reroughening rate is very fast, and is almost saturated after only 10 seconds exposure to dDF. Not enough data in shorter time ranges has been taken as yet to make clear conclusions about the roughening rate. However, another feature of these spectra should be pointed out here. A clear H_2 monohydride peak is observable even after only 5 seconds in the dDF. From this result, we can conclude that there is significant H contamination in the dDF solution. A quantitative measurement of this contamination can be made by comparing H_2 intensities after 1 minute in dHF only (I_o) and after 1 minute in dHF followed by dDF exposure until the H/D exchange rate is saturated (I_s), about 60 minutes. Such a measurement shows that I_s = 18% I_o. This agrees very well with Kinoshita's results which showed I_s = 15% I_o for similar experiments using Si(100) as the substrate.[22] However, these hydrogen contaminations are much higher than the calculated value of 2.4% due to the hydrogen in the HF and H_2O that is mixed with the D_2O to make the dDF solution. The reason for this discrepancy has not yet been determined. It is because of this H contamination, we believe, that 100% H/D exchange was not possible by simple exposure of an H terminated Si surface to dDF as was reported above.

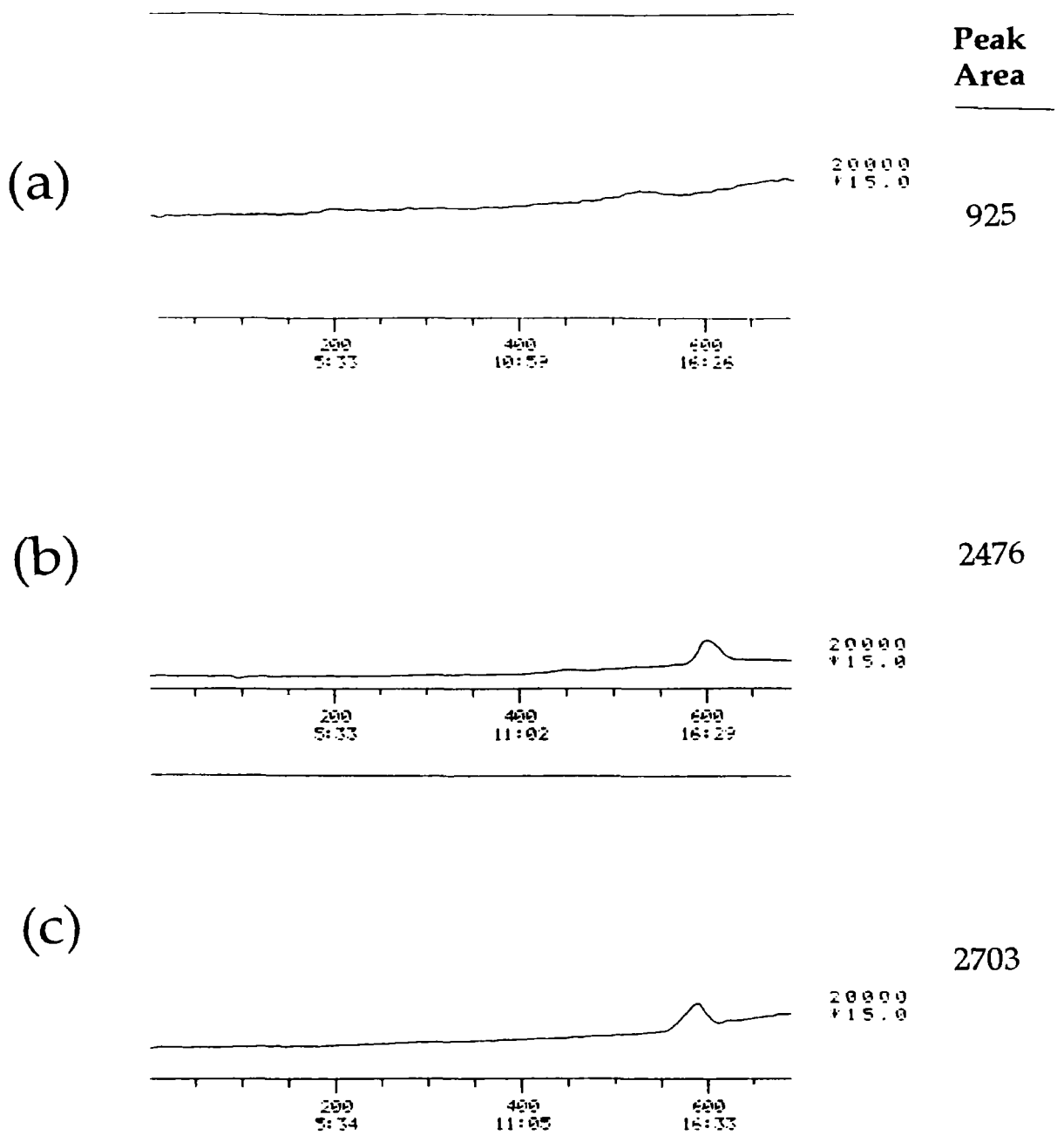


Figure 23: TDS spectra of SiO of Si(111) samples prepared by: (A) rough, deuterium termination -> exposure to clean room air at room temperature for 6 days, (B) rough, deuterium termination -> baked in a clean oven at 325°C for 1 hour, (C) APM treatment only. To the side of each of the spectra is the area of the SiO peaks as measured by the TDS software in arbitrary units.

3-8 References for TDS Intensity

While the intensity units measured by the TDS software are arbitrary, we can get a general idea of relative peak intensities by comparing different peak areas measured by the TDS software. Figure 23 shows SiO TDS spectra for samples (A) exposed to air at room temperature for about 6 days, (B) baked in a clean oven at T=325°C for 1 hour, and (C) APM treated only. The respective areal peak intensities of each are given next to the spectra. It should be noted that the APM treated samples and those baked at 325°C have very similar intensities (about 8% difference). While there is some disagreement as to the exact thickness of the oxide film left on the Si surface after APM treatment, a generally accepted value is about 12 Å. Therefore, we can assume that the thickness of the oxide layers in our study is between 0 and 12Å. Furthermore, if we assume that the TDS intensity varies linearly with oxide thickness, we can convert the measured SiO peak intensity to oxide layer thickness in Å. While the validity of the last assumption may be questionable since we only have two data points, it still allows us to get a general idea of the oxide layer thickness as a function of baking temperature.

3-9 Time Dependence of SiO Formation

While the principle study of this report is the temperature dependence of SiO formation, the time dependence allows us to compare our data with that of previous reports in the thin film range. Figure 24 shows oxide thickness as a function of time in an oxidizing environment at T=200°C. The curve has been fitted to the data by a form of the well known

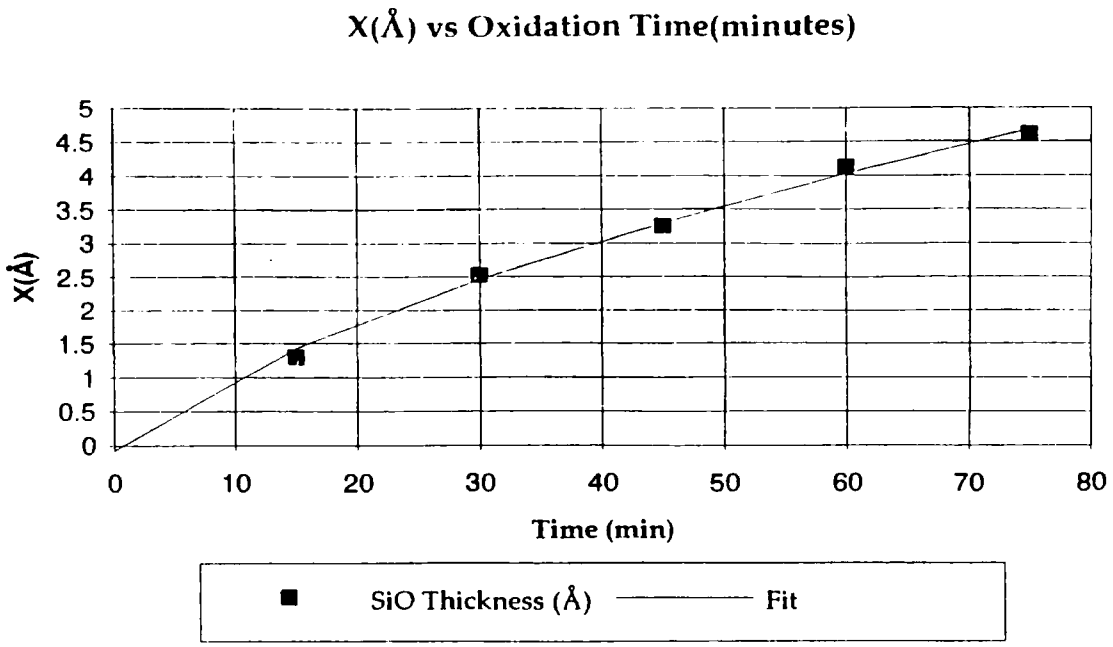


Figure 24: Plot of oxide thickness vs oxidation time at a constant temperature of 200°C. The curve is given by the linear-parabolic relationship given by eqn. (1) with the parameters: a_o=0.62, a_1=7.56, and a_2=1.79.

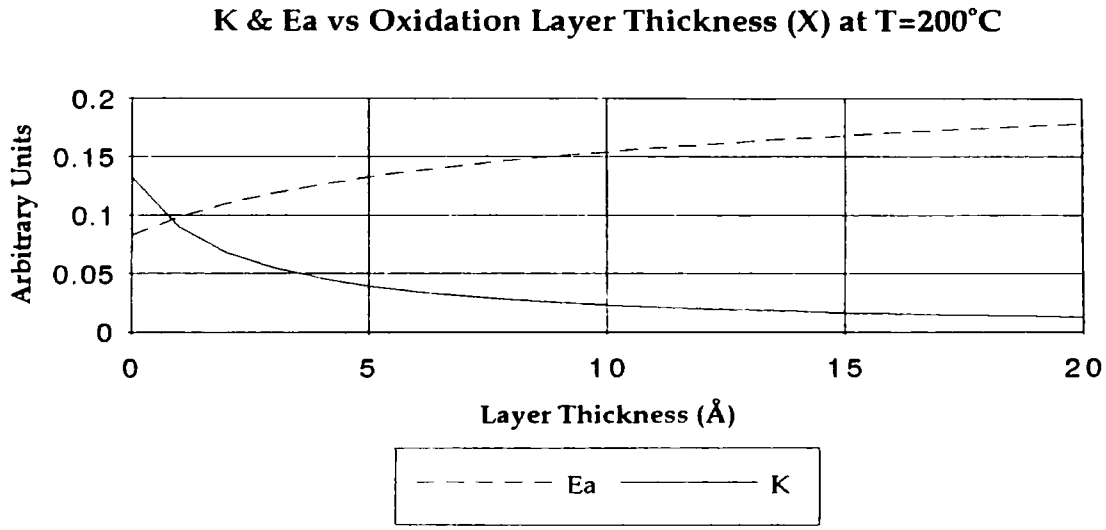


Figure 25: Plot of the activation energy E_a and reaction rate K as functions of oxide thickness. Parameters are the same as those in Figure 15.

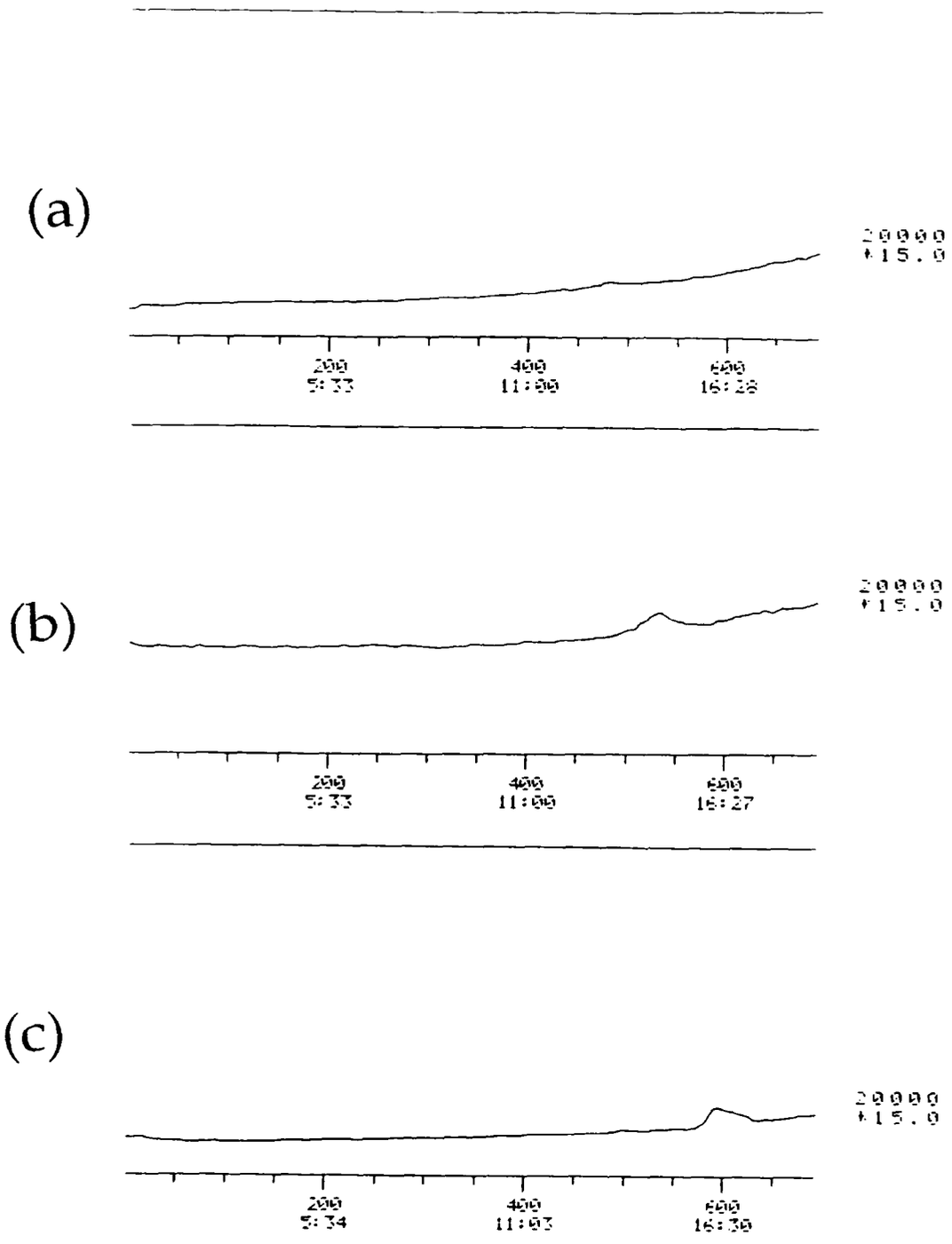


Figure 26a: TDS spectra of SiO of Si(111) samples prepared by: rough, deuterium termination -> baked in a clean oven for 1 hour at (A) 100°C, (B) 200°C, (C) 300°C.

linear-parabolic law reported by Deal and Grove.[23] This curve is given by the second order polynomial:

$$t = a_0 + a_1 x + a_2 x^2 \qquad (1)$$

where t is the oxidation time, X is the oxidation layer thickness, and the a_i's are constants that depend on temperature and have been adjusted to fit the data. As Fig. 24 shows, this curve fits our data quite well with $a_0 = 0.62$, $a_1 = 7.56$, and $a_2 = 1.79$. The oxidation velocity can be found by differentiating Eq. (1) with respect to X and inverting:

$$\frac{dx}{dt} = K(x) = \frac{1}{a_1 + 2a_2 x} \qquad (2).$$

Many studies have shown that the measured oxidation rate deviates from that predicted by the Deal-Grove model, and expressed by equation (2), in the thin film ranges.[24,25,26,31,32] The measured values are consistently higher than those predicted. Several different kinetics models have been proposed which have better agreement with the experimental data in both the thin and thick film range. These, however, have rather complicated expressions for X(t) and K(t or X). So, since our data agrees well with the Deal-Grove model, we will assume this type of kinetics for the first analysis. Fig. 25 shows a plot of K(X) using the parameters from the fit of Eq. (1) to our data.

3-10 Temperature Dependence of SiO

For these oxidation experiments, the exposure time of the samples to the oxidizing environment was held constant while the temperature of that environment is varied. Figure 26A shows some examples of SiO TDS

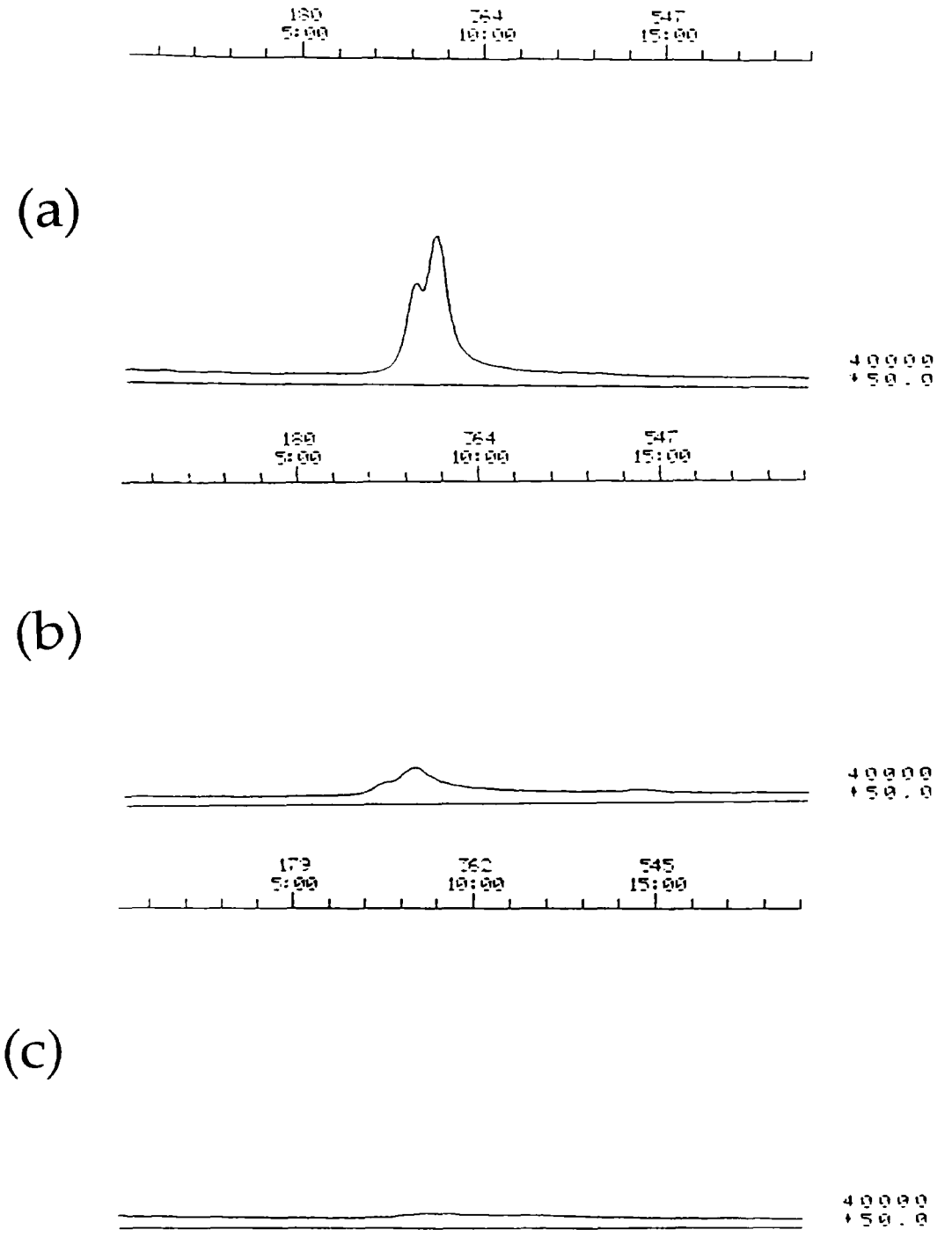


Figure 26b: TDS spectra of D_2 of Si(111) samples prepared by: rough, deuterium termination -> baked in a clean oven for 1 hour at (A) 100°C, (B) 200°C, (C) 300°C.

spectra for rough surface samples oxidized at various temperatures. As is expected, the SiO intensity grows with temperature. Figure 27 shows a plot of the oxide thickness X as a function of temperature for both smooth and rough D terminated surfaces. Several rather surprising features were observed. First, for temperatures above about 150°C, both smooth and rough samples show a linear temperature dependence. Also, a linear regression analysis of the data in this region shows that slopes of the best fit lines agree to within 1%. This indicates that above 150°C, both smooth and rough samples oxidize in a similar fashion. The difference between the samples seems to lie in the lower temperature range. Extrapolations of the best fit lines show that the threshold temperature, that is the temperature at which oxidation begins during the 1 hour baking time, of the rough surface is 65°C and for the smooth surface is 115°C. So the threshold temperature is shifted by 50°C. Some reports have suggested that oxidation of the Si surface has an initial stable state in which O is inserted into the Si backbonds between the first and second layers of the crystal (e.g. at a step site or defect) rather than the H(D) terminated dangling bonds.[27] The shift in the threshold oxidation temperature observed in our data may be explained by this proposal since the atomically rough surface has many more step sites and defects than the smooth surface.

If the oxidation reaction rate K follows a normal Arrhenius type temperature dependence, then

$$K(T) = Ae^{\frac{-E_a}{kT}} \tag{3}$$

where E_a is the activation energy, A is the preexponential factor, and k is Boltzmann's constant. If E_a is constant, then X(T) should show an

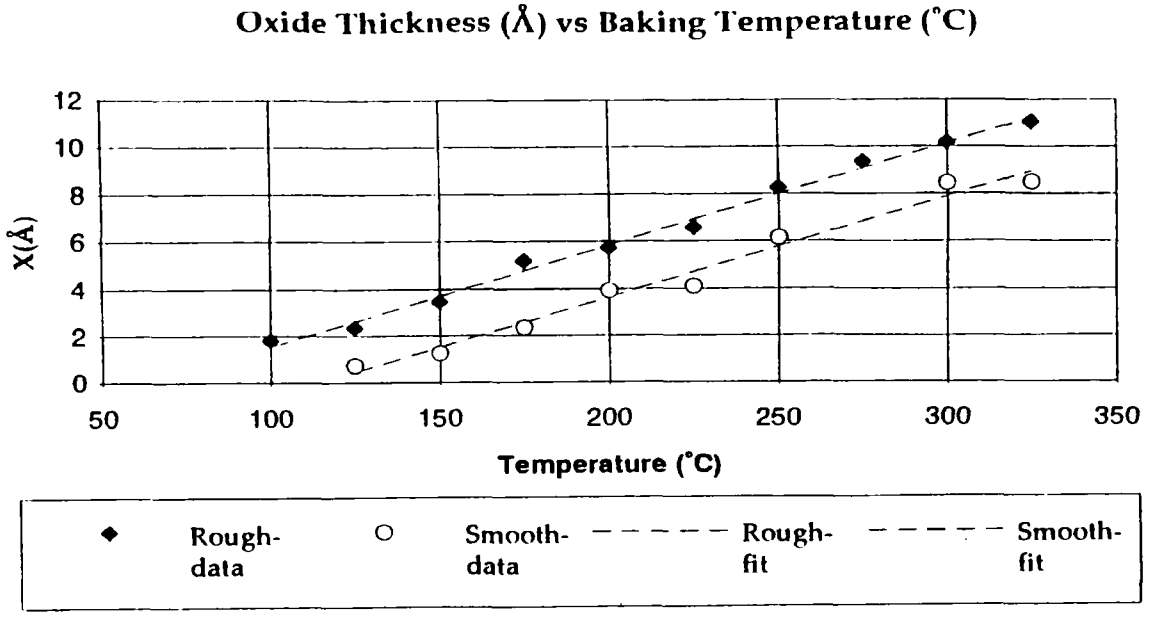


Figure 27: Plot of the oxide thickness X (Å) vs oxidizing temperature (°C).

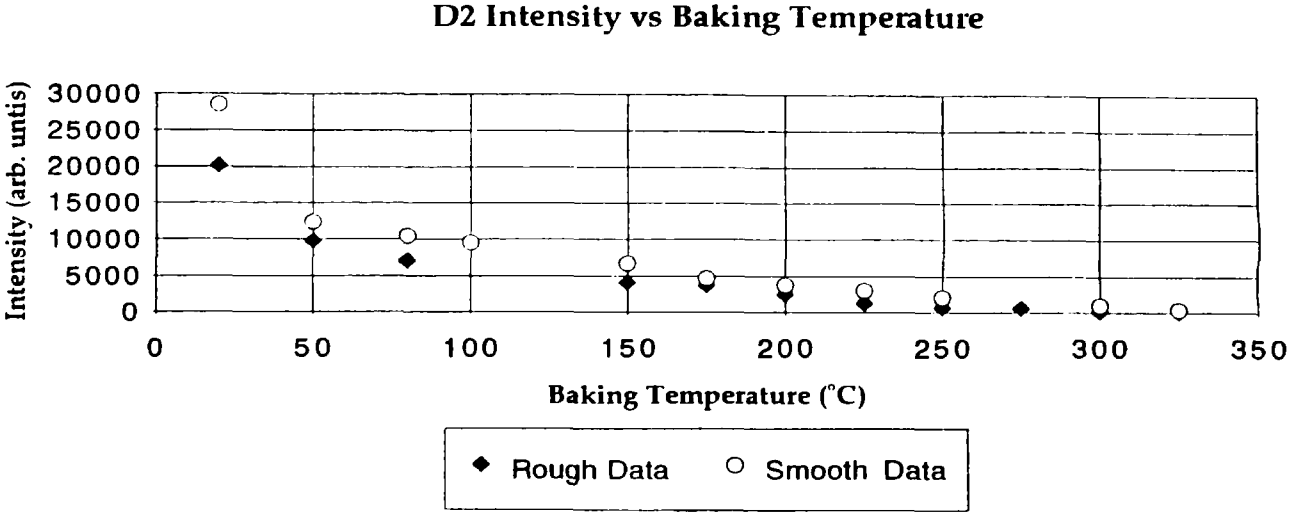


Figure 28: Plot of D_2 intensity (arb. units) vs oxidizing temperature (°C).

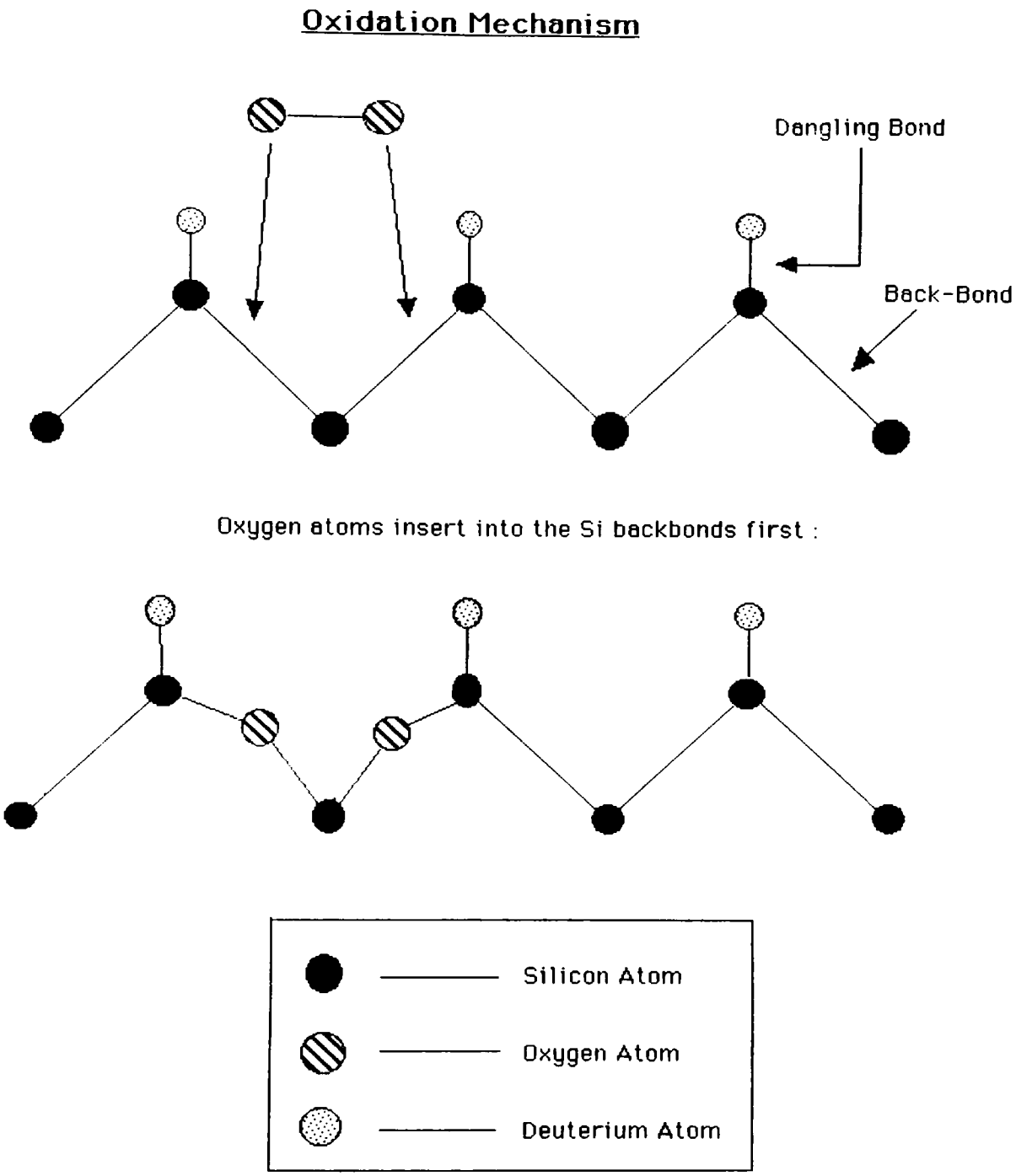


Figure 29: Oxidation mechanism that seems to be supported by our data. Oxygen initially inserts itself into the Si backbond, and only later attacks the H(D) terminated dangling bonds.

exponential increase. However, as Fig. 27 shows, X(T) has a linear relationship with temperature in the range observed (150 < T < 325°C). This linear behavior of X would seem to suggest that K(T) does not follow the normal Arrhenius type behavior of Eq. (3). However, Eq. (3) assumes that the activation energy is a constant parameter. This has been a valid assumption for relatively thick films (> 100Å) , but some reports have shown that the activation energy becomes dependent on oxide thickness in the thin film range (< 100 Å) .[21,23,28] Since the maximum thickness of our oxide layer is only about 12-15 Å, it seems that our assumption of a constant E_a is invalid, and the linear behavior of X can be explained by postulating that E_a is a function of oxide thickness.

If Eq. (2) is assumed to give the correct K(X), then from Eqs. (2) and (3), we can solve for $E_a(X)$. This gives the result:

$$E_a(x) = -kT \ln\left(\frac{1}{a_1 + 2a_2 x}\right) \qquad (4)$$

if the preexponential term A is considered to be a constant and absorbed by the arbitrary units. $E_a(X)$ is plotted with K(X) in Fig. 25 for a constant temperature of T = 200°C. This curve is in qualitative agreement with the previous reports that the activation energy increases with oxide thickness in the thin film range and approaches some constant value for large oxide thickness X. Obviously, further and more careful analysis is needed and should be the subject of future work.

3-11 Deuterium Temperature Dependence

Also of interest in this study is the deuterium intensity as a function of oxidizing environment temperature. Figure 26B shows some examples of

D_2 TDS spectra for rough Si(111) samples at various temperatures. It is clear that the D atom terminating the Si dangling bonds are almost completely removed as the temperature reaches about 300°C. Figure 28 shows a plot of the D_2 intensity as a function of oxidizing environment temperature. As can be seen, in the range of T=22 to 325°C, the intensity roughly behaves as an exponential decay for both smooth and rough surfaces. By plotting the natural log of the D_2 areal peak intensity, both sets of data become quite linear indicating an exponential behavior. The rough and smooth data have approximately the same decay rate, but the smooth data is shifted about 30° higher. This is in qualitative agreement with our SiO data. The difference in intensity temperature dependence between SiO and D_2, linear and exponential respectively, and the lack of a threshold temperature for D_2, would seem to indicate that the mechanism for the oxidation of H(D) terminated Si is not a simple exchange process. This would also agree with the earlier mechanisms proposed in which O inserts into the Si backbond initially and only later attacks the surface dangling bonds terminated by the D atoms. (see Fig. 29) Once again, a more careful analysis of the data is needed before any firm conclusions about the initial oxidation mechanisms can be made.

Chapter 4: Conclusions & Proposed Future Research

There are several important conclusions that can be drawn from these experiments. The first of which is that we have found a relatively simple process for creating an atomically smooth and deuterium terminated Si(111) surface. Terminating the surface with deuterium rather than hydrogen has no particular advantages in semiconductor device production, but it does have a real advantage for studying the surface science of silicon in that atmospheric deuterium is much lower than hydrogen. This makes observations of the surface phenomenon much easier, and can lead to advances in semiconductor devices. The process for making such a sample is as follows:

1) chemically clean the sample by APM treatment

2) remove any native oxides by soaking the sample in dDF (1.6% in heavy water) for 10 minutes

3) boil sample in very dilute DF (0.005% in heavy water).

The second conclusion is that our experimental data are consistent with the initial oxidation model in which the O is first inserted into the Si backbond between the first and second monolayers. This is opposed to the model in which the O replaces the D that terminates the Si dangling bond. Future experiments should be made to provide further evidence for the backbond model.

The third conclusion, and perhaps the one most important to semiconductor device production, is that the onset of oxidation is retarded by smoothing the Si(111) surface. Our results clearly show a 50°C shift in the

oxidation threshold temperature between the smooth and rough surfaces. Therefore, by boiling the silicon wafers for 10 minutes, the formation of native oxides can be greatly reduced during the time before the wafers are thermally oxidized.

In order to achieve a better understanding of the oxidation process in this ultra-thin film range, more experiments and analysis are needed. For instance, more oxidation time dependence experiments, perhaps at different oxidation environment temperatures, would provide a better idea of how the kinetics in this oxidation layer thickness range deviates from the Deal-Grove law. This would also provide a way in which to compare these temperature dependence results with more time dependence experiments which are quite prevalent in the literature. Another question which has been left unanswered here is why boiling the Si(111) samples in heavy water should leave an oxidation layer when the same process using normal water does not. While this phenomenon may not have great importance for the semiconductor device industry, it is a very puzzling question which deserves some academic interest. Lastly, it would be interesting to increase the oxidizing environment temperature range up to thermal oxidation temperatures (~1000°C) to see if the linear dependence holds throughout this range. Also, by using several different detection methods, such as TDS in conjunction with photoelectron spectroscopy, the accuracy of the experiments could be increased and more direct comparisons with previous studies could be made.

Appendix

Basic Theory of the Quadrupole Mass Spectrometer

The mass spectrometer has been an invaluable tool for scientists since its inception by J.J. Thomson about eighty years ago. These tools allow us to measure the masses and relative abundance of atoms or molecules in a sample. Because of its wide range of applications, the mass spectrometer has undergone many modifications and adaptations as more specific applications were desired. All mass spectrometers work on the same basic principle. They produce an ion beam from some sample, accelerate the ions by an electric field, separate the ions according to their mass-to-charge ratio (m/e) with the help of a magnetic field, and produce an output signal which is a measure of the abundance of the species present. There has been a great deal of research in creating different methods of carrying out each of these steps, which has resulted on a myriad of different kinds of mass spectrometers. The type of spectrometer used in this research is the quadrupole mass spectrometer (QMS). In this section we present a general discussion of what a QMS is and the principles by which it works.

The idea of the QMS was originated by W. Paul and H. Steinwedel in 1953[29], and falls into the category of dynamic mass spectrometers which is defined by E.W. Blauth as one in which some fundamental component of the system, like the electric or magnetic field strength, or ion movement, is time dependent.[30] The QMS falls under a sub category of the path-stability type. These types of spectrometers use a radio-frequency (rf) field to vary the trajectory of the ions so that only those with a narrow range of masses will stay within a prescribed region of space and be detected. In other words, the

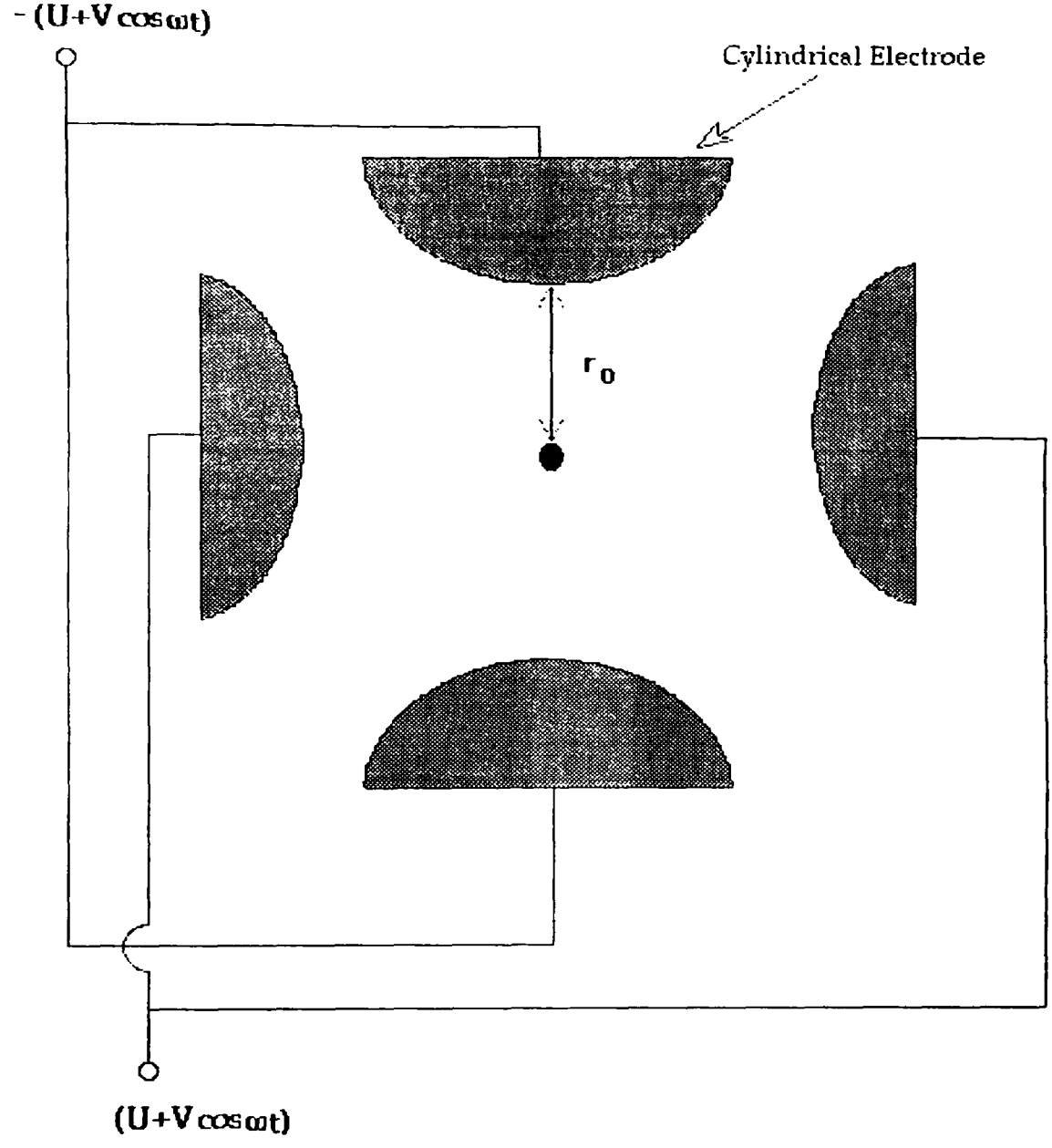


Figure 30: This figure shows a schematic of the electrodes which produce the quadrupole field in the QMS where r_0 is the distance from any of the electrodes to the center of the array. The axis of symmetry of the array points straight out of the paper as does the accelerating field which pushes the ions along the apparatus to the detector.

QMS acts like a mass filter ; it only lets ions with a narrow mass range through to the detector.

The QMS achieves this task with electric fields only. Four long cylinders are arranged in a square with the centers separated by a distance of $2r_o$ where r_o is the inside radius of the array of cylinders. (see Fig. 30). The rods opposite each other are electrically connected with an rf potential, V cos ωt, and a dc potential U. The dc potentials are equal in magnitude but opposite in sign, while the rf potentials are phase shifted by 180°. This arrangement produces an electric quadrupole field with a potential of

$$\varphi(x,y,z,t) = (U + V\cos\omega t)\frac{(x^2 - y^2)}{r_o^2}$$

with the z-axis parallel to the cylinder axes and located at the center of the array.[31] As can be seen from the formula, the potential is zero along the lines $|x| = |y|$.

The ions are accelerated along the z axis from a source at one end of the cylinders toward an electron multiplier by a voltage V_{acc}. If the potential are set to a ratio of $U/V = 0.17$ then a particular rf frequency will allow ions of a specific m/e ratio to pass through the filter.[32] This ratio is obtained by a stability diagram which is a phase space plot of two dimensionless parameters involving U, V, and M. These parameters, usually denoted by α and q, are incorporated into a system of differential equations, which define an ions trajectory in a quadrupole field. By adjusting these parameters, stable solutions of the motion equations can be found. This value for the U/V ratio is found by picking parameter values in this stable region of the stability diagram. All ions of a different mass will have unstable trajectories and collide with one of the cylinders before they reach the end of the filter.

The abundance of the one species allowed through is measured using an electron multiplier whose signal is amplified and measured by some recording device, a desk top computer running spectroscopy software in this case.

If a mass spectrometer could only measure the abundance of one species, it would not be a very useful instrument. The equation relating the mass number in amu of the species allowed through the mass filter and rf frequency is $M=0.136V/(r_o^2f^2)$, where M is in amu, V is in volts, r_o is in cm, and f is in megahertz. So scanning various mass numbers can be done with the QMS in two ways, either by varying V or f. The first option is the most common method. However it should be noted that the U/V ratio must remain constant, so U must be changed accordingly. The less commonly used method is to vary the frequency while keeping the voltages constant. In our experiments, we wanted to observe the mass spectra of different species as the silicon was heated. To do this, the QMS was set to quickly cycle through up to nine different V and U magnitudes with a period of about 1.5 seconds. In other words, each species was measured about every 1.5 seconds. Since the temperature ramping took about 18 minutes, this gave us good resolution.

An important measure of a mass spectrometers performance is its resolution. Resolution is a measure of the spectrometer's ability to separate ions of slightly different masses. It can be measured using the formula $R=M/\Delta M$, where M is the mass of the peak in amu and ΔM is the width of the peak at 5% of the peak height. There is always some trade off for higher resolution. In the QMS, better resolution comes at the expense of transmission percentage, or the percent of ions being measured that make it

through the mass filter to the electron multiplier. In order to adjust the resolution of the QMS, the U/V ratio must be adjusted. The ratio of 0.17 given above provides an adequate balance between resolution and transmission. In order to achieve a very high resolution of 100, the transmission is about 10%, while 100% transmission yields a resolution of about 35.

The most important advantages of the quadrupole mass spectrometer over other types of spectrometers are 1) it is not dependent on the initial energy of the incident ion beam, 2) a relatively high transmission percentage, and 3) fast mass scanning. The last advantage is probably the most important for this work since we were usually observing nine species at one time.

References

[1]Chih-Tang Sah, *Proceedings of the IEEE,* 76, No. 10, October 1988. Unless otherwise noted, all information in this section was obtained from this article.
[2]Raymond Serway, *Physics for Scientists and Engineers, Second Edition,* Saunders College Publishing, 1986.
[3]G. K. Teal, M. Sparks, and E. Buehler, Physical Review, 81, 637 (1951).
[4] W. Shockley and W. T. Read Jr., Physical Review, 87, 835 (1952).
[5]M. M. Atalla, M. Tannenbaum, and E.J. Scheibner, Bell Systems Technical Journal, 38, no. 3, 749-783.
[6]R. H. Dennard, F.H. Gaensslen, H.N. Yu, L. Rideout, E. Bassous, A.R. LeBlanc, J. Solid-State Circuits, vol. SC-9, pp. 256-268, May 1973.
[7]B.E.A. Saleh, M.C. Teich; *Fundamentals of Photonics,* p.433-434, John Wiley and Sons, Inc., 1991.
[8]Raymond Serway, *Physics for Scientists and Engineers, Second Edition,* Saunders College Publishing, 1986.
[9]Ibid.
[10]A.S. Grove, *Physics and Technology of Semiconductor Devices,* p.208-211, John Wiley and Sons, Inc., 1967.
[11]Ibid., p.318.
[12]Ibid., p.320.
[13]Ibid., p.327.
[14]J. M. Feldman, *The Physics and Circuit Properties of Transistors,,* 532, John Wiley and Sons, Inc., 1972.
[15] T. Sakurai and H.D. Hagstrum; Physical Review B 12(12), 5349 (1975).
[16] Watanabe, Shigeno, Nakayama, Ito; Extended Abstracts of the 1991 International Conference on Solid State Devices and Materials, Yokohama, Japan, 1991, pp. 502-504.
[17] C.H. Mak, B.G. Koehler, and S.M. George; Surface Science 208 L42 (1989).
[18] G.S. Higashi, Y.J. Chabal, G.W. Trucks, K. Raghavachari; Applied Physics Letters 56 (7) 656 (1990).
[19] E. Yablonovitch, D.L. Allara, C.C. Chang, T. Gmitter, T.B. Bright; Physical Review Letters 57 (2), 249 (1986).
[20] C.M. Greenlief, S.M. Gates, P.A. Holbert; Chemical Physics Letters,159, 202 (1989).
[21] B.G. Koehler, C.H. Mak, D.A. Arthur, P.A. Coon, S.M. George; J. Chemical Physics 89(3), 1709 (1988).
[22] Kinoshita; NEC Internal Technical Report; 2-10-92.
[23] B.E. Deal and A.S. Grove; J. Applied Physics, 36, 3770 (1965).
[24] Y. Kamigaki and Y. Itoh; J. of Applied Physics, 48 (7), 2891 (1977).
[25] Y.Z. Lu and Y.C. Cheng; J. of Applied Physics, 56, 1608 (1984).
[26]Kamohara and Y. Kamigaki, Extended Abstracts of the 22nd Conference on Solid State Devices and Materials, Sendai, Japan, 1990, pp. 1087-1090.
[27] U. Hofer, P. Morgen, W. Wurth, and E. Umbach; Physical Review B, 40 (2) , 1130 (1989).
[28] Massoud, Plummer, and Irene; J. Electrochemical Society,2685 (1985).
[29]W. Paul and H. Steinwedel, Z. Naturforsch., 8a, 448 (1953).
[30]E. W. Blauth, *Dynamic Mass Spectrometers,* Elsevier, Amsterdam, Netherlands, 1966.
[31]John Roboz, *Introduction to Mass Spectrometry, Instruments and Techniques,* p.106, John Wiley and Sons, Inc.,1968.
[32]Ibid., p.107.